Computerised diagnosis of malaria.

F. Boray Tek

School of Life Sciences

COMPUTERISED DIAGNOSIS OF MALARIA

F. Boray Tek

A thesis submitted in partial fulfilment of the

requirements of the University of Westminster

for the degree of Doctor of Philosophy

September 2007

COMPUTERISED DIAGNOSIS OF MALARIA

F. Boray Tek, Ph.D.
University of Westminster 2007

Malaria is a very serious infectious disease caused by a peripheral blood parasite of the genus *Plasmodium*. According to the World Health Organization statistics, it causes more than 1 million deaths arising from approximately 300-500 million infections every year. Microscopy examination of the blood films is currently "the gold standard" for the malaria diagnosis. However, this method has not been changed since the late 19th century.

Manual microscopy diagnosis of the blood film is subjective and time consuming. The aim of this work has been to investigate the potential of performing this task without human intervention and to provide an objective, reliable, and efficient tool to do so.

Diagnosis of malaria infection is enhanced using a chemical process called staining. Staining highlights parasites but also some other regular blood components. Microscopy diagnosis requires examination of stained blood films and detection of parasites in the case of an infection. An automatic diagnosis tool must provide a method for detecting parasites by differentiating between parasites and regular peripheral blood components.

Earlier studies on the topic provided many auxiliary tools to extract the stained objects by applying or developing techniques based on general image processing and mathematical morphology. They also investigated the possibility of identifying life-stages of parasites using heuristic rules. However, the most important problem -parasite detection- was not addressed. Hence, diagnosis of any reliable sort and consequently a fully automated one could not be performed.

This work investigates the application of machine learning and statistical pattern recognition perspectives to the problem. It proposes a method for parasite detection based on K-Nearest Neighbour classification. Parasite detection is the fundamental function of automated diagnosis.

This work extends the proposed parasite detection method to provide species and life-stage recognition which are also necessary functions. Species recognition enables identification of the *Plasmodium* species for appropriate treatment. Life-stage recognition aids assessing the parasite development in infected blood.

This work proposes a successful method for colour normalisation of thin blood film images. This enables processing of the images regardless of their acquisition characteristics. The normalisation method is not specific for *Plasmodium* therefore can be used as a pre-process in other peripheral thin blood film analysis methods.

It also proposes two new methods in the field of mathematical morphology: the Minimum Area Watershed Transform and the Radon Transform-based Marker Extraction algorithms. These are then applied to the thin blood film image segmentation problem.

This work provides a complete set of solutions for thin blood film malaria diagnosis. However, some of the solutions such as colour normalisation or stained pixel/object detection are not specific to the characteristics of *Plasmodium*, or its appearance in a particular set of images. Hence, they can be incorporated in any thin blood film analysis problem. In addition, the classification methodologies for parasite detection, species and life-stage recognition are generalised solutions. This is a very positive attribute of the overall methodology that it is not limited to *Plasmodium* but it is equally and effectively applicable to diagnosis of other blood parasites. Moreover, it is anticipated that an automated system based on the proposed methodology will be expandable and thus will be able to perform differential diagnosis, provided that the same features are representative and discriminative.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisors, Assoc. Prof. Andrew G. Dempster and Prof. İzzet Kale for their excellent guidance and support throughout the course of this research. I would also like to thank Dr. Ediz Çetin for his support and guidance in almost every problem I faced.

Thanks to my examiners, Dr. Andrzej Tarczynski (internal) and Prof. Yong Yan (external) for reading the thesis thoroughly and for their valuable comments and corrections, which gave this thesis its final state.

I would like to thank Dr. Martin Parry of Department of Biology, University of Westminster (London, UK) for organising the short course "Update in Blood Parasites" and for his efforts which made it possible for me to attend the course. I would like to extend my thanks to the course instructors Monika Kettlehut (Hospital for Tropical Diseases, London, UK) and Claire Rogers (London School of Hygiene and Tropical Medicine) for teaching me diagnosis of malaria thoroughly and for their patience with my extra questions. This course has completely changed my understanding of malaria diagnosis and thus gave the main structure to this thesis.

I would also like to thank Dr. William Jarra (National Institute of Medical Research, London, UK) for providing us with many sample images and for kindly sparing time from his busy schedule to use the parasite marking software.

I would like to thank to Dr. Carolina Galmes for reading my thesis and for her valuable comments.

Thanks to Mustafa Taşkaldıran, Süleyman Demirsoy, İzzet Özçelik, Eric Qu, Ali Telli, Kostadin Tzvetkov for their useful discussions and excellent friendship.

Thanks to my flatmates Asanka and Dawya for their patience, support, and excellent friendship especially during the past 6 months of the thesis.

I am very grateful to my brother Buğra Tek who financially supported me during the past difficult couple of months.

Finally, I would like to thank my girlfriend Mercedes Rodriguez and my parents Vahide and Tekiz Tek for their love and encouragement, which made finishing this thesis possible.

ASSOCIATED PUBLICATIONS

1. F. B. Tek and A. Dempster and I. Kale, Malaria parasite detection in peripheral blood images, *British Machine Vision Conference*, Sep 2006, Edinburgh, UK.

2. F. B. Tek and A. Dempster and I. Kale, Malaria parasite detection in peripheral blood images, *Medical Image Understanding and Analysis Conference*, July 2006, Manchester, UK.

3. F. B. Tek and A. G. Dempster and I. Kale, A colour normalization method for Giemsa-stained blood cell images, *Signal Processing and Applications* (Turkish title, Sinyal Isleme ve Uygulamalari), Apr 2006, Antalya, Turkey.

4. F. B. Tek and A. G. Dempster and I. Kale, Blood cell segmentation using minimum area watershed and circle radon transformations, *International Symposium on Mathematical Morphology*, Apr 2005, Paris, France.

5. F. B. Tek and A. G. Dempster and I. Kale, Noise sensitivity of watershed segmentation for different connectivity: experimental study, *IEE Short Letters*, 40(21), pp 1332-1333, 2004.

Submitted

- F. B. Tek and A. G. Dempster and I. Kale, Parasite Detection and Identification for Automated Thin Blood Film Malaria Diagnosis, *Medical & Biological Engineering & Computing Journal, Springer.*

- F. B. Tek and A. G. Dempster and I. Kale, Computer Vision For The Microscopy Diagnosis of Malaria, *IEEE Transactions on Information Technology in Biomedicine.*

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

BPNN Back Propagation Neural Network

CCD Charge-Coupled Device

CDM Computerised Diagnosis of Malaria

CMOS Complementary Metal-Oxide-Semiconductor

FLD Fisher Linear Discriminant

FN False Negative

FP False Positive

HSV Hue Saturation Value

KNN K-Nearest Neighbour Rule

MM Mathematical Morphology

MAWT Minimum Area Watershed Transform

NPV Negative Prediction Value

PCA Principal Component Analysis

PPV Positive Prediction Value

RBC Red Blood Cell

RDT Rapid Diagnosis Tests

RGB Red Green Blue

ROC Receiver Operating Characteristic

SE Sensitivity

SP Specificity

TN True Negative

TP True Positive

UoW University of Westminster

WBC White Blood Cell

LIST OF SYMBOLS

Morphological Operation Symbols

$\ominus$ Binary Erosion, e.g. $A \ominus S$: A is eroded by S

ϵ Grey level erosion, e.g. $\epsilon_S(f)$ is erosion of image f by structuring element S

$\oplus$ Binary Dilation, e.g. $A \oplus S$: A is dilated by S

δ Grey level dilation, e.g. $\delta_S(f)$ is dilation of image f by structuring element S

γ Opening, e.g. $\gamma_S(f)$ is opening of image f by structuring element S

ϕ Closing, e.g. $\phi_S(f)$ is closing of image f by structuring element S

γ^a_λ Area Opening, e.g. $\gamma^a_\lambda(f)$ is area opening of image f by area threshold λ

ϕ^a Area Closing, e.g. $\phi^a_\lambda(f)$ is area closing of image f by area threshold λ

WTH White Top-Hat, i.e. Top-hat by opening

BTH Black Top-Hat, i.e. Top-hat by closing

WTH^a White Area Top-Hat, i.e. Top-hat by area opening

BTH^a Black Area Top-Hat, i.e. Top-hat by area closing

R^δ Reconstruction by Infinite Dilation, e.g. $R^\delta_g(f)$ is reconstruction of f by dilation of g

$T_{[a,b]}$ Grey level image single threshold

$TT_{[a,b,c,d]}$ Grey level image double threshold by reconstruction

$RMAX$ Regional Maxima

$RMIN$ Regional Minima

IZ Influence Zone

General Symbols

$.*$ Pixel-wise multiplication

$*$ Matrix Convolution

$|||$ Grey level histogram

Chapter 1

Introduction

Computerised Diagnosis of Malaria (CDM) should not be thought of as a newly discovered medical diagnosis technique. Instead, it tries to translate the medical expertise and knowledge of microscopy diagnosis of malaria to a computer platform. By this translation one can provide constant medical expertise and reliability wherever the technique is employed.

To provide some information and set the scene about the problem of malaria diagnosis, the first three sections of this chapter explain the malaria disease, malarial parasite and current diagnosis techniques. Then, a general view of computerised diagnosis of malaria is presented. Sections 1.5, 1.6, and 1.7 present research aims, contributions, and an outline of this thesis. The final two sections provide brief explanations about the methodology used in this study.

1.1 Malaria Disease

Malaria is a very serious infectious disease caused by a protozoan parasite of the genus *Plasmodium*. Despite it being preventable and curable, it causes more than 1 million deaths arising from approximately 300-500 million infections every year. While in total 107 countries have areas at risk from malaria, it is mainly active in Sub-Saharan Africa because the environmental conditions are suitable for mosquitoes and the poor socio-economic conditions make access to health and prevention resources difficult [1]. The estimated annual cost of malaria for African countries is US$12 billion [2]. The World Health Organization (WHO) estimates the annual minimum cost for the effective control of malaria is around US$3.2 billion [1]. However, in reality only a small fraction of this is available to fight against malaria. The Roll Back Malaria Partnership initiative aims to increase annual global spending on malaria prevention, treatment, and research

to US$4.1 billion by 2010 [3].

Many information resources about malaria is readily available on the internet, the major ones being the WHO [4], Roll Back Malaria [3], Center for Disease Control and Prevention (CDC) [5], and the Malaria Journal [6].

1.2 *Plasmodium*: The Malarial Parasite

Malaria is transmitted by infected female Anopheles mosquitoes which carry Plasmodium sporozoites in their salivary glands. The complete life cycle has separate development stages in the human and mosquito body. When an infected mosquito feeds on a person's blood, the sporozoites enter the blood stream and move to the liver where they multiply asexually for a period. Then they produce merozoites to enter the peripheral blood stream again to invade Red Blood Cells (RBCs) [5, 7]. Within the RBC the parasite either grow until it reaches a mature form and break the cell to release more merozoites into the blood stream to invade new RBCs, or alternatively, it may grow to reach a sexual form named gametocyte (female and male) to be in taken by a mosquito where it sexually reproduces to produce sporozoites. The illustrations of the life cycle of malaria can be found in various online sources (e.g. [4, 5]).

1.2.1 Parasite Species and Life-Stages

Genus *Plasmodium* has 4 species that can cause human infection. These are *Plasmodium Falciparum*, *Plasmodium Vivax*, *Plasmodium Ovale*, and *Plasmodium Malariae*. There are other species which infect animals [7, 8]. During the life cycle in peripheral blood, different species may be observable in the five different life-stages which are generally morphologically distinguishable (Figures 1.1, 1.2, 1.3, and 1.4). Some of the life-stages may be observable together. Life-stages common to all species are merozoite, ring, trophozite, schizont (which are all asexual) and gametocytes which are the sexual forms. Asexual stages can be further sub-categorised by their maturity

and gametocytes are generally sub-categorised by the genre (female or male). Figure 1.1 illustrates various life-stages of *Plasmodium Falciparum* that are observable in peripheral human blood. Different species usually have different morphology in the same life-stage category while in some life-stages it may be hard to distinguish between species. For example, the first 10 illustrations of *P. Falciparum* in Figure 1.1, the first 5 illustrations of *P. Vivax* in Figure 1.2, the first 5 illustrations of *P. Ovale* in Figure 1.3 and the first 5 illustrations of *P. Malariae* in Figure 1.4 correspond to ring stages of the species which are not easily distinguishable. The illustrations presented here are taken from CDC's malaria website [5] with written permission and was first published in [7]. There are other illustrations which can be used as guides for species and life-stage morphologies [9, 10].

1.3 Diagnosis Techniques

There are two classical techniques used for malaria diagnosis, and there is a new group of Rapid Diagnosis Tests (RDT), which have become available recently. A widely used approach is clinical diagnosis based on the symptoms. The clinical examiner makes a decision based on the symptoms such as high fever, headache, severe chills and body pain [9]. Since these symptoms are not specific to malaria, the accuracy of the clinical diagnosis can be very low [11]. This approach is often preferred in the field where access to the second classic approach (microscopy diagnosis) is not available or not efficient because it requires a specially trained examiner. Provided that the sample is well prepared and the examiner is trained, microscopy diagnosis is very reliable. It is sensitive and specific, which enables the determination of the type of the malaria infection [11].

Diagnosis using a microscope requires special training and considerable expertise. Various studies have aimed to assess the accuracy of the microscopy diagnosis. It has been shown that what is known as "field microscopy" is not an effective malaria

Figure 1.1: *Plasmodium Falciparum* life-stages in thin films. 1: Normal red cell, 2-10: ring-stage trophozoites, 11-18: Trophozoites, 19-26: Schizonts (26 is a ruptured schizont); 27-28: Macrogametocytes (female); 29-30: Microgametocytes (male). Reproduced with permission [5].

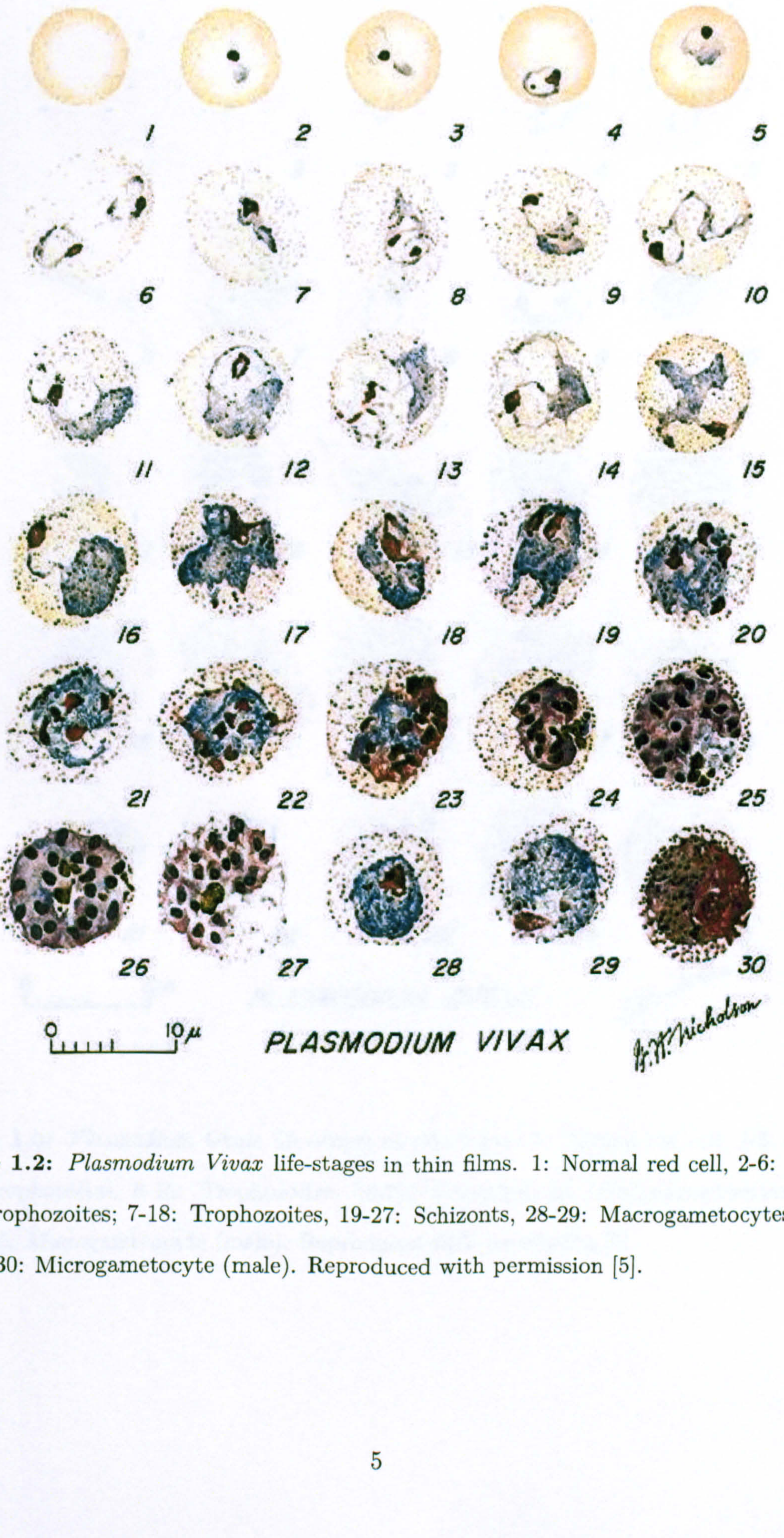


Figure 1.2: *Plasmodium Vivax* life-stages in thin films. 1: Normal red cell, 2-6: Ring stage trophozoites; 7-18: Trophozoites, 19-27: Schizonts, 28-29: Macrogametocytes (female), 30: Microgametocyte (male). Reproduced with permission [5].

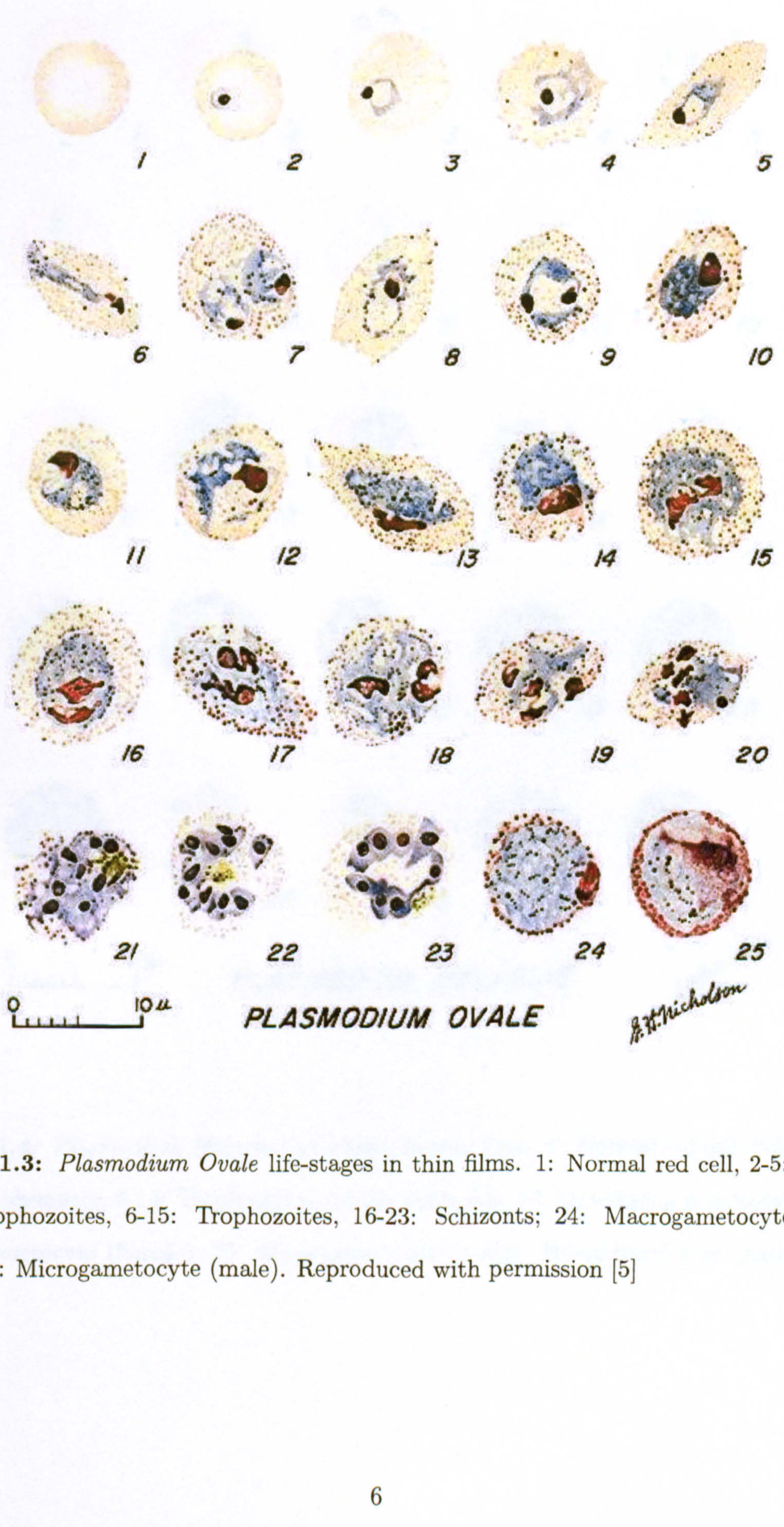


Figure 1.3: *Plasmodium Ovale* life-stages in thin films. 1: Normal red cell, 2-5: Ring stage trophozoites, 6-15: Trophozoites, 16-23: Schizonts; 24: Macrogametocytes (female),25: Microgametocyte (male). Reproduced with permission [5]

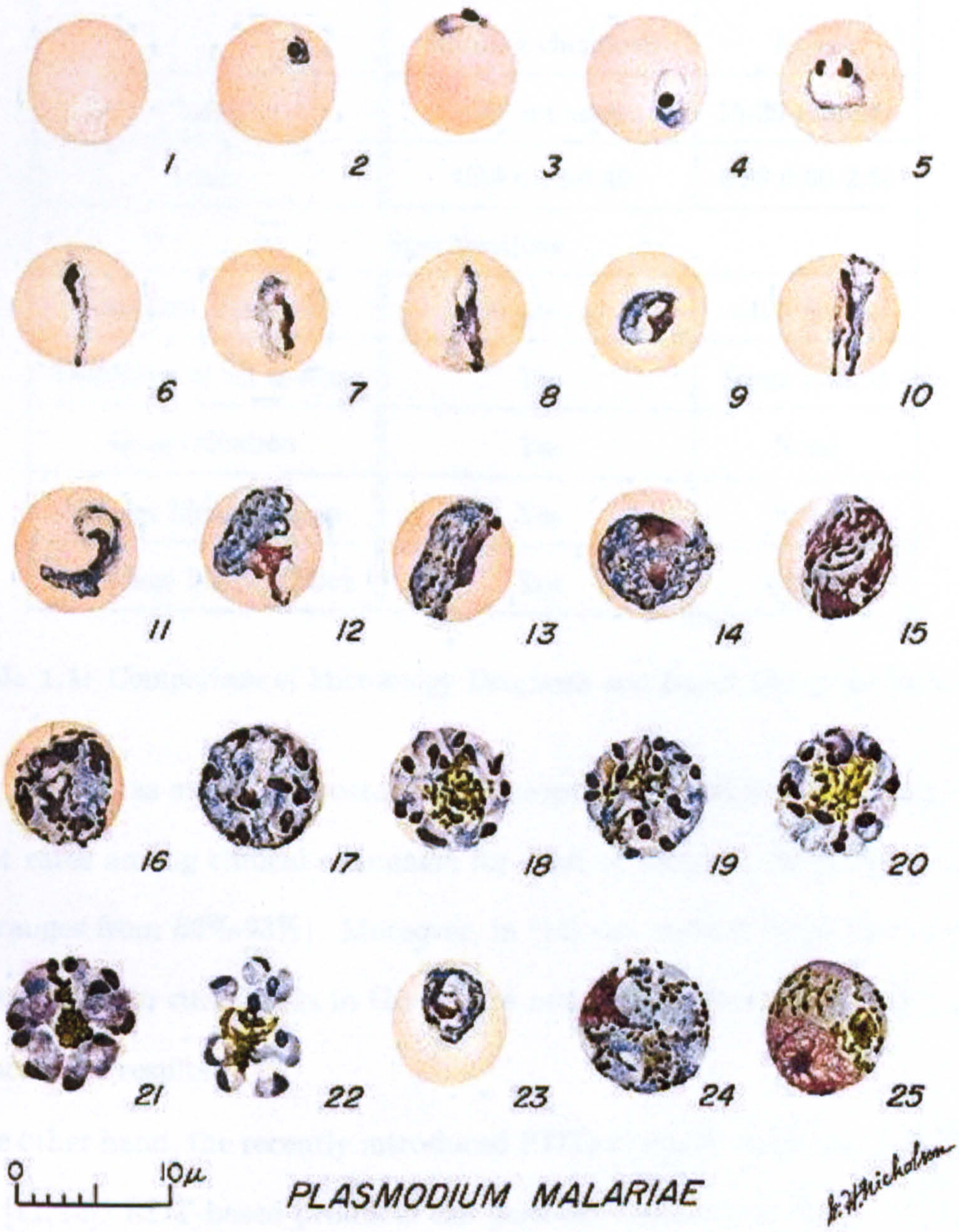


Figure 1.4: *Plasmodium Malariae* life-stages in thin films. 1: Normal red cell, 2-5: Ring stage trophozoites; 6-13: Trophozoites, 14-22: Schizonts, 23: Developing gametocyte, 24: Macrogametocyte (female), 25: Microgametocyte (male). Reproduced with permission [5].

	Microscopy	RDTs
Requirements	Electricity (optional) Special Training Staining chemicals	None Basic Training None
Time	~60 minutes	15-20 minutes
Cost	US$ 0.12-0.40	US$ 0.60-2.50
Specifications		
Detection Threshold	50 $par/\mu l$	~100 $par/\mu l$
Detection of all species	Yes	Some brands
Quantification	Yes	None
Species Identification	Yes	None
Life-Stage Identification	Yes	None

Table 1.1: Comparison of Microscopy Diagnosis and Rapid Diagnosis Tests [11]

screening method as expert laboratory microscopy [12]. A field study shows that the agreement rates among clinical examiners for a set of samples are below expectations [13] (i.e. ranges from 66%-93%). Moreover, in [14] the authors state that most of the clinical examiners in rural areas in Ghana are not trained and experienced enough to produce accurate results.

On the other hand, the recently introduced RDTs provide an alternative for malaria diagnosis [11, 15]. RDT-based products can perform diagnosis in about 15-20 minutes and require no special training, equipment or electricity. Detection sensitivities of RDTs are comparable to microscopic diagnosis for higher levels of parasitemia (i.e. parasite density). However, they do not provide quantitative results and the specificity for differentiating the species. Currently the cost of one examination is several times greater than the cost of microscopy examination. Table 1.1 provides a general comparison of microscopy diagnosis to RDTs.

There are methods other than those mentioned above, such as molecular analysis

Polymerase Chain Reaction (PCR) which is more accurate than microscopy but is more expensive, needs a specialised laboratory and takes days to produce a result [5,16]. The PCR is known as the most sensitive parasite detection method; it can detect 5 or less parasites/μl of blood [17]. However, it is difficult to test the sensitivity of the PCR because the reference (i.e. expert manual microscopy) is not as sensitive as the test. A recent comparison shows that the PCR is clearly more accurate than expert laboratory microscopy for malaria at parasite densities $\geq 500/\mu l$ [18]. Other techniques which rely on fluorescence microscopy need more expensive microscopes than the conventional ones [19, 20]. More reviews on the diagnosis techniques can be found in [17, 21–23].

1.3.1 Microscopy Diagnosis

Microscopy diagnosis of malaria is performed by examining Giemsa stained peripheral blood slides using a conventional light microscope. The slides for examination can be prepared in two different forms: thin and thick films. In thick films 3 drops of blood from the patient are concentrated in $1cm^2$ area on the slide while in thin films 1 drop of blood is spread to form a single layer of cells. Additionally, thin films are fixed with methanol to preserve RBCs whereas in thick films the blood is de-hemoglobinised in water which results in the only observable components being White Blood Cells (WBCs) and parasites (or stained structures). Because of the concentration in a small area, the thick film examination provides more sensitivity for the parasite detection. However, because the RBCs are destroyed, species differentiation is more difficult than the thin film examination. The average sensitivity thresholds for thick and thin films are reported as 50 parasites/μl (0.0001%) and 500 parasites/μl (0.001%) of blood, respectively (1μl of a normal blood sample contains ~5 million RBCs). According to a recent WHO report [17], an experienced laboratory examiner can detect parasites on thick films in concentrations as low as 5-10 parasites per μl of blood. However, this depends on the skill of the technician and the time spent for the examination. For

	P. Falciparum	P. Vivax	P. Ovale	P. Malariae
Size	Not enlarged	Enlarged	Enlarged	Not Enlarged
Shape	Round crescent gametocyte	Round or oval	Round, oval amoeboid	Round
Dots	Large red spots	Small red dots	Small red dots	Few tiny dots

Table 1.2: Some Rules for Species Identification

example, a study in the UK for the sensitivity of routine malaria diagnosis showed the sensitivity is on average 500 parasites/μl [24].

If malaria parasites are detected, species identification is *essential* for appropriate treatment [25]. There are published rules which take into account some visual properties such as the changes of the infected RBC shape or some specific dots which enables differentiation between different species (see Table 1.2) [5, 9]. However, there is no guarantee that every parasite example will exhibit the morphological characteristics. Figure 1.5 demonstrates this with 3 different ring stage trophozoites of *P. Vivax* where the small red dots (Schuffner dots), and RBC enlargement is clearly observable in R1 and R2, yet it is difficult to say the same for the third (R3). If the gametocyte G2 was not there, R3 can be difficult to specify. For these reasons species recognition is a complicated task, which may require examination of other parasites in the same specimen. Parasites can be more distinctive in the more mature life-stages compared to the ring life-stage. For example, the identification of *P. Falciparum* can be simpler if some gametocytes exist and are observable in the smear. Additionally, mixed infections (i.e. more than one species) are not rare which implies a thorough examination is required to be certain.

The life-stage categorisation is mainly concerned with maturity of the parasites or establishing whether it is a gametocyte or not. Identification of parasite life-stage aids assessing the parasite development in infected blood. However, since the parasite is a biological and growing form, maturity has no clearly divided boundaries. It is possible

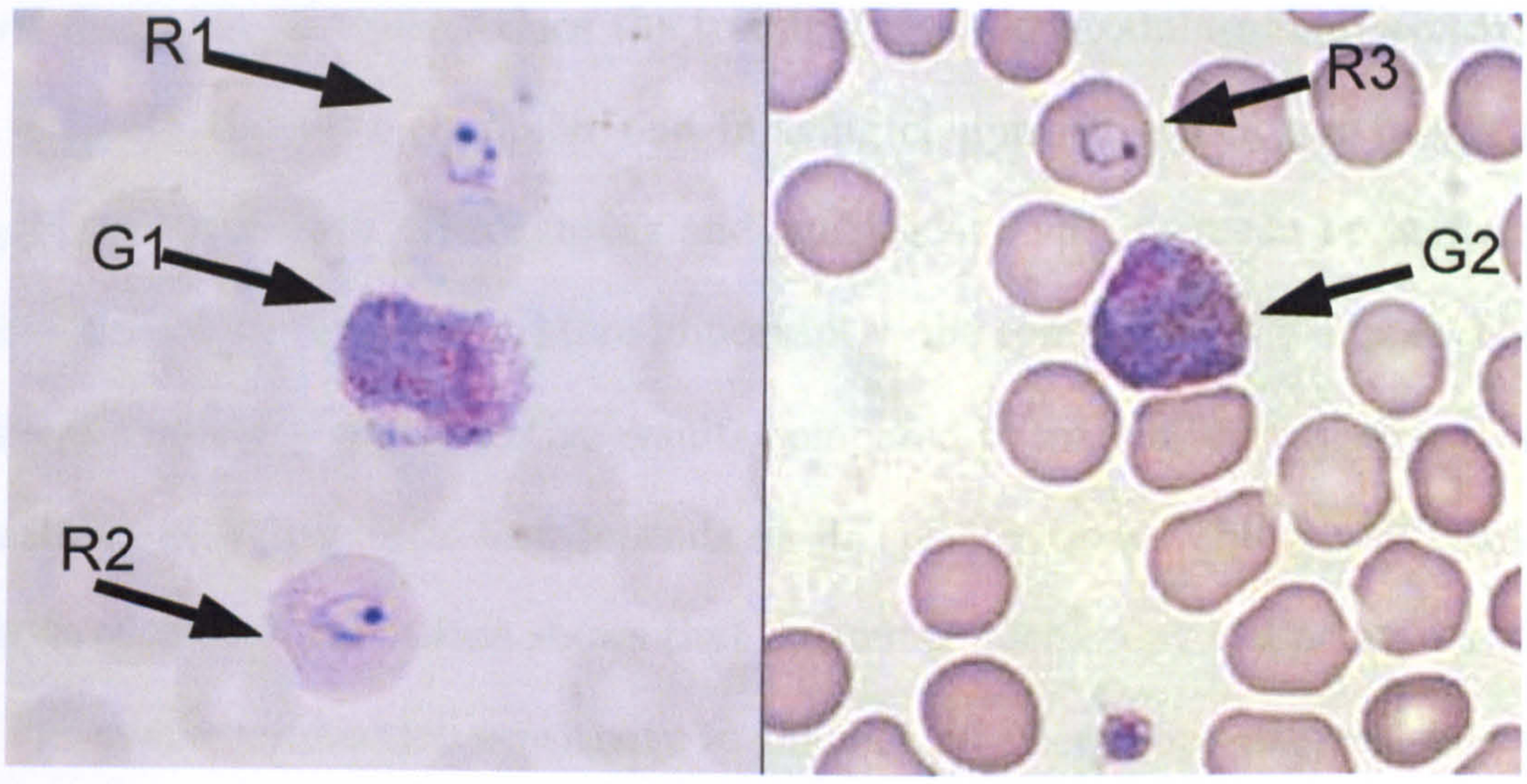


Figure 1.5: *P. Vivax* parasites R1, R2, R3 are ring stage trophozoites, G1 and G2 are gametocytes. Image on the left is from University of Westminster malaria image set, the image on the right is from [10].

to be confused with a transition stage parasite, e.g. a mature ring-early trophozoite (Figure 1.1 9-14), a mature trophozoite-early schizont (Figure 1.2 18-19), or a mature trophozoite-early gametocyte (Figure 1.4 9-23) parasite.

1.4 Computerised Diagnosis

Computerised diagnosis of malaria is a microscopy diagnosis technique. It can be used as an aid or a complete automated diagnosis technique, which replaces the manual microscopy examination. Obviously, its ability to replace an expert depends on the accuracy of its diagnosis performance. A computerised diagnosis system can be used in various areas such as research, clinical diagnosis, evaluation of the treatments, blood screening or in any place where *Plasmodium* should be observed, counted or linked to other clinical data.

Before starting a discussion of the methodology, we can discuss possible advantages and disadvantages of the computerised microscopy diagnosis. As listed in Table 1.3, the requirements for computerised microscopy diagnosis would be similar to those for

manual diagnosis. It could reduce the training needs for examiner significantly; however, it would require a computer and imaging equipment, which can makes it less accessible in rural areas. The imaging and analysis of a specimen can be performed in constant time with computers. More importantly, the computerised diagnosis can provide more consistent and objective results compared to manual microscopy. However, the usability of a diagnostic test depends on its specifications. The overall framework that is developed in this thesis shows that computer microscopy can achieve a comparable Plasmodium detection sensitivity to the manual microscopy diagnosis, detect all four Plasmodium species, and provide species and life-stage information.

A computerised microscopy diagnosis technique can be designed by understanding the microscopy diagnosis expertise and representing it by specifically tailored image processing/analysis, and pattern recognition algorithms. However, in order to be practical, the complete system will require some additional functions such as positioning of the slides, report generation; and a database of the diagnosis records and the demographic data of the patients. However, the study and research in this thesis is only concerned with the problem from the computer vision point of view, which is seen as the most essential part. A basic computer vision system to perform the fully automated diagnosis has to provide solutions for the sub-problems listed below:

- *Image Acquisition*: This is an essential task, which requires interfaces between the camera, the microscope and the computer. There are expensive pre-built systems commercially available, which already have this functionality. However, considering the relationship of the disease with regions with poor resources, a better alternative can be a modular low-cost configuration, which is a combination of a consumer camera with an ordinary microscope. However, an image acquired with an unknown setup may have various characteristics, which may result in very different image appearances. Therefore, the ideal system should be designed to cope with images acquired under various uncontrolled conditions and perform

	Manual microscopy	Computer Microscopy
Requirements	Electricity (optional) Special Training Staining chemicals	Yes Basic Training Same+Computer+Camera
Time	Subjective	Constant
Accuracy	Subjective	Constant&Consistent
Cost	US$ 0.12-0.40	Similar
Specifications		
Detection Threshold Thick Films	50 *par/µl*	–*
Detection Threshold Thin Films	500 *par/µl*	~690* *par/µl*
Detection of all species	Yes	Yes
Quantification	Yes	Yes
Species Identification	Yes	Yes
Life-Stage Identification	Yes	Yes

Table 1.3: Comparison of Manual and Computerised Microscopy Diagnosis Requirements and Specifications.*This thesis studies thin film analysis. However, due to constrictions of data, the study could not include per-specimen tests; ~690 *par/µl* thin film sensitivity threshold is based on a hypothetical comparison (see Section 8.4).

the same processing on them. This can be achieved using either of two different strategies: providing a calibration procedure for the acquisition setup or a robust method for the pre-processing of the images to reduce the effects. This thesis provides a detailed methodology for the latter which is explained in Chapters 3 and 5.

- *Field Selection*: A single blood slide can have thousands of separate fields (images) at the required magnification for the diagnosis. Moreover, the blood is not homogeneously distributed among these fields. It may be necessary (e.g. for segmentation) to automate the selection of the fields to be analysed. This requires full control of the microscope functions such as focus and x-y table movement. If we assume the microscope provided is fully controllable, an algorithm can be developed according to some criteria (e.g. thickness, number of cells, spreading) to decide suitability of fields for succeeding processes. This is not a complicated problem and there is a study in the literature that proposes a solution based on mathematical morphology [26]. In this thesis, we propose a method, which can process images regardless of the field density, which removes the necessity of employing a field selection procedure (Chapter 6).

- *Segmentation*: Segmentation of the image aims to separate blood cells into separate regions. With segmentation, the blood cells in a thin film can be partitioned and labelled uniquely for further processes. Since thick films do not have resolvable RBCs, this type of segmentation is not possible or required. Segmentation of thin film images can be useful because usually the degree of infection (parasitemia) is calculated as the ratio of the number of infected cells to the healthy ones. However, segmentation is not the only solution for estimating the number of red blood cells or parasitemia. In addition, the counting of RBCs may be unnecessary if the process can provide the parasite count against the number of white blood cells or a fixed blood volume. This thesis provides a more detailed

discussion on the thin blood film image segmentation problem in Chapter 4.

- *Parasite Detection*: To enable the observation of parasites, a blood sample is stained prior to the examination. While almost all of the components are affected by the staining, the parasites, WBCs and platelets are significantly highlighted. Detection of parasites requires the detection of stained pixels. Then the detected stained pixels should be further processed to define the objects. These stained objects should then be classified to determine if they are parasites. This classification requires the definition of the parasite together with the definition of the non-parasite. By the parasite we are referring to *Plasmodium* species which infect human (P. Falciparum, P. Vivax, P. Ovale, and P. Malariae), however it must be noted that there are other peripheral blood parasites which are responsible for other diseases such as Babesiosis, Filariasis, and Leishmaniasis, Trypanosomiasis [27] or animal malaria [7, 8]. By the non-parasite we are referring to the WBCs, platelets and artefacts, however there may be other anomalies, which can enlarge the class definition. We will discuss the properties of parasite and non-parasite objects and parasite detection problem in more detail in Chapters 7 and 8.

- *Identification of parasite species and life-stages*: The difficulties of species identification (or recognition) were mentioned earlier (Section 1.3.1). Species recognition, which is required for the detected parasites, is a further pattern classification problem that has four target classes: *P. Falciparum*, *P. Vivax*, *P. Ovale*, *P. Malariae*. The distinctive properties of different species were given in Table 1.2. Species recognition can be performed according to these rules. However, since certain species have distinctive morphologies in certain life-stages, it may be possible to combine two separate problems of species and life-stage recognition. For example, the classification can be performed using sixteen target classes (i.e. four species x four life-stages) instead of four. We will discuss these in Chapter 9.

- *Calculating parasitemia and reporting*: A complete diagnosis should first report if parasites are observed in the blood slide (smear). Then, if the blood is infected, the ratio of the infected cells to healthy ones and species (and life-stages) should be reported. The ratio of infected cells to healthy ones calculated over a standard blood volume is called parasitemia. It is a common practice for experts to count parasitemia for only *P. Falciparum* since other species do not reach intense populations. For computer diagnosis, provided all previous tasks are complete, producing a parasitemia report for all malaria species is straightforward. Parasitemia in a thin film is calculated as the ratio of the total number of the ring, trophozite and schizont stage parasites to the number of red blood cells (commonly per an average of 1000 RBCs). However, depending on the efficacy and speed of the analysis, other measures can be calculated such as the ratio of the number of parasites to the WBCs or the complete categorised parasite and WBC count.

1.5 Research Aims and Methodology

As discussed above, the computerised diagnosis problem consists of many sub-problems. Due to the size of the problem, the scope of this thesis is focused on achieving only the following listed aims:

1. Image acquisition

2. Colour normalisation of images

3. Segmentation of images

4. *Plasmodium* detection in thin films

5. *Plasmodium* species and life-stage recognition in thin films

There are many different paradigms for image analysis and pattern recognition in computer vision, which can be utilised to achieve these goals. However, we extensively use mathematical morphology for image analysis and statistical learning based approaches for pattern analysis. Section 1.8 provides a general description of mathematical morphology and gives definition of some morphological operators, which are extensively used in this research. Section 1.9 provides a brief definition of general pattern recognition terms (e.g. feature, classifier, and training). More detailed information can be found in [28] and [29] for mathematical morphology and in [30–33] for statistical pattern recognition.

1.6 Contributions

The automated diagnosis of malaria has been the subject of two earlier research studies at the University of Westminster. First, Dr. Cecilia Di Ruberto [34] and later on Dr. K.N.R. Mohano Rao [35] have studied the thin film examination. These studies focused mostly on processing of blood cell images to provide the basis for parasite analysis. They have also developed some rule-based methods for life-stage analysis. In Di Ruberto's work [34] the presence of regular blood components in images such as WBCs and platelets were briefly addressed. In Rao's work [35] the presence of regular blood components was not considered at all. Every stained object in the thin film images was considered as parasites. For automated diagnosis, this approach would diagnose every possible sample (and consequently the patient) as infected even though they may be completely healthy.

Automated microscopy diagnosis of malaria essentially requires a function, which can differentiate between regular blood components (also artefacts) and parasites. The significance of this thesis study is that it defines the computerised "diagnosis of malaria" directly and thoroughly. It proposes a novel method to perform this fundamental function (i.e. parasite detection).

Species and life-stage analysis must be performed *only* on the detected (i.e. identified) parasites. The study in this thesis proposes a method to perform these secondary functions.

This work also reinvestigates some of the auxiliary functions (e.g. segmentation and stained object detection) which were also investigated during the previous research. This was necessary to allow processing of all the fields of thin films regardless of their cell concentration (i.e. thickness).

The main contributions and related publications resulted from this research are listed below:

- This work proposes a novel method for parasite detection based on the K-Nearest Neighbour classification rule. The proposed method was presented in British Machine Vision Conference (Edinburgh, 2006) [36], and Medical Image Understanding and Analysis Conference (Manchester, 2006) [37] .

- This work proposes and compares three alternative classification schemes for the species and life-stage recognition problems based on the K-Nearest Neighbour classification rule. A comprehensive article including the species and life-stage recognition results is currently under preparation.

- This work proposes a novel method for colour normalisation of thin blood film images. This enables processing of the images regardless of their acquisition characteristics. The normalisation method is not specific for *Plasmodium* therefore can be used as a pre-processing stage in any peripheral thin blood film analysis method. The method was presented by the author in 15th Signal Processing and Applications Conference (Antalya, 2006) [39].

- This work contributes to the field of mathematical morphology by proposing the minimum area watershed transform and Radon transform based marker extraction algorithms. These were applied to the thin blood film image segmentation

problem and were presented at the International Symposium of Mathematical Morphology (Paris, 2005) [40].

In addition to the above contributions, in this study, many of the proposed methods were comprehensively evaluated and validated. To perform various experiments a significant number of images (peripheral thin blood film images) were acquired from a setup, which was physically built during this study.

This study approaches the computerised diagnosis of malaria problem from a pattern classification perspective. The set of solutions proposed in this thesis provides a well-defined basis for computerised diagnosis of malaria. It proposes generalised solutions for the sub-problems which avoid heuristics and rule based algorithms. As an outcome, it is possible to design a generalised diagnosis tool simultaneously covering other peripheral blood parasites since the methodology does not depend on the characteristics of *Plasmodium*, or its appearance in a particular set of images. This is very important because a diagnosis tool would be more powerful if it can identify other parasites. This is discussed and emphasised throughout the thesis.

1.7 Outline of Thesis

The rest of this chapter (Sections 1.8 and 1.9) give brief explanations of widely used mathematical morphology tools and pattern classification notions. Readers who are familiar may prefer to skip them and return for reference when necessary.

Chapter 2 presents a literature review of the studies that are directly related to the diagnosis of malaria. Additionally, it discusses some other studies, which are not directly related to malaria but are related to the analysis of the blood cell images or have completely different subjects but share some common techniques with this thesis.

Chapter 3 explains our research related to the image acquisition and data collection. It discusses the problem of uneven illumination and highlights the possible solutions.

Chapter 4 presents a novel method for the segmentation of the blood cell images

and discusses the problems associated with segmentation.

Chapter 5 presents a novel method for the colour normalisation of the thin blood film images.

Chapter 6 explains our approach to the stained pixel detection problem and presents a new solution for stained pixel detection using a Bayesian pixel classifier. Additionally, it proposes an algorithm for the post-processing of the stained pixels to extract the stained objects.

Chapter 7 presents our methodology for malaria diagnosis from the statistical pattern recognition perspective.

Chapter 8 presents our experimental study for parasite detection. It includes a comprehensive set of experiments pointing to the features, classifier details and scale-invariance.

Chapter 9 presents our experimental study of the species and life-stage recognition problems. The species and life-stage recognition tasks are seen as extensions to parasite detection. However, the study considers the three tasks as both consecutive and joint. It proposes and compares different classification schemes for the tasks.

Chapter 10 gives a summary of the thesis, marks the achievements and contributions to the field and provides conclusions and some future directions. It discusses the study in terms of fully automated malaria diagnosis.

1.8 Mathematical Morphology

Mathematical morphology (MM) is a theory concerned with morphology or spatial analysis of images. It is based on set theory and lattice algebra [41]. It provides many powerful techniques for image analysis which are particularly useful for general blob image analysis tasks such as blood cell image analysis. The suitability of MM to such tasks stems from the advantage of set (object) level manipulation of images. Many of the functions provided within mathematical morphology are based on nonlinear

transformations in which the output is given as the relation between input set and a transforming set which is defined by an attribute. This attribute can be based on an increasing measure such as width, size, basic shape (square, circle etc.) or can be non-increasing measure such as circularity, elongation, or contrast. Some key definitions and functions, which are extensively used in this study, are presented here. The theoretical details and more information can be found in texts [28, 29].

- **Connectivity**

  Connectivity in a digital image can be explained with graphs. A connected graph is the graph in which any two of its vertices can be linked by a path. For example, if two sample points $[(x1, y1), (x2, y2)]$ in a digital image are taken as vertices: these vertices are said to be 4-connected if $|x1 - x2| + |y1 - y2| = 1$, whereas they are said to be 8-connected if $max(|x1 - x2|, |y1 - y2|) = 1$. The connectivity definition is required to define the neighbourhood of a specific coordinate in the image plane. Almost all the MM operations can produce different results according to the different connectivity definition. We have shown in [38] that in watershed segmentation the difference can be significant for noisy images. In this thesis, all the MM operations use 8-connected connectivity definition.

- **Structuring Element**

  A structuring element can be defined as a set which modifies a morphological operation's behaviour with a defined attribute (e.g. shape, size). In its simplest form, a structuring element will consist of a pattern specified by discrete coordinates relative to an origin. By choosing the coordinates, specific shapes can be formed. More complex structuring elements can include defined values for each specified coordinate. In some operations the structuring elements can be implicitly defined in an adaptable manner according to an attribute instead of predefined set of coordinates.

- **Erosion and Dilation**

  These are the basic operations of MM which are used in derivation of the higher level operations. Suppose X denotes the set of coordinates x corresponding to the input image A, S is the set of coordinates defined by the structuring element, and Sx denotes the translation of S so that its origin is at x. Then the erosion of A by S is defined as the set of points x such that Sx is a subset of X [29]:

$$A \ominus S = \{x : S_x \subseteq X\} \tag{1.1}$$

This can also be formulated by intersection of set of translations (i.e. Minkowski subtraction), that the translation is determined by S can be defined as:

$$A \ominus S = \bigcap_{s \in S} X_s \tag{1.2}$$

The erosion is a translation, order invariant and increasing operation [29]. The opposite (dual) of erosion is dilation which is defined as set of all points x such that the intersection of Sx with X is non-empty [29]:

$$A \oplus S = \left\{x : \hat{S}_x \cap X \neq \emptyset\right\} \tag{1.3}$$

Here, again the definition can be made by union of set of translations (i.e. Minkowski addition), that the translations being determined by the S can be defined as:

$$A \oplus S = \bigcup_{s \in S} X_s \tag{1.4}$$

The extension of morphological operations to grey scale images is done by perceiving the image as a topological surface. Binary set operations intersection and union are replaced by the minimum and maximum operations respectively. For a grey scale image (f), grey scale erosion and dilation operations can be defined as in [29], respectively.

$$\epsilon_S(f) = \min_{s \in S} f(x + s) \tag{1.5}$$

$$\delta_S(f) = \max_{s \in S} f(x + s) \tag{1.6}$$

- **Opening and Closing**

  The combination of dilation and erosion operations produces two new operations. These are opening and closing depending on the order which operation is applied first. Opening is erosion followed by dilation:

$$\gamma_S(f) = \delta_S(\epsilon_S(f)) \tag{1.7}$$

  And closing is the opposite (dual) of the opening:

$$\phi_S(f) = \epsilon_S(\delta_S(f)) \tag{1.8}$$

  Opening and closing are also dual operations and they are translation invariant, increasing and idempotent. While opening is anti-extensive, closing is an extensive operation [29].

- **Area opening and Area closing**

  Area opening is an opening operation which defines a locally adaptable structuring element based on an area attribute instead of providing a fixed width structuring element [42]. The area opening operation removes an object if its area is below a defined threshold λ. Again area closing is defined as the dual operation of area opening. In grey scale images area opening removes a plateau and area closing fills a hole if its area is below a defined threshold. Area opening for binary images is defined as follows [42]:

$$\gamma_\lambda^a(X) = \{x \in X | Area(C_x(X)) \geq \lambda\} \tag{1.9}$$

  where $C_x(X)$ the connected opening or simply connected component of X. Area closing is defined by duality [42]:

$$\phi_\lambda^a(X) = [\gamma_\lambda^a(X^c)]^c \tag{1.10}$$

  where X^c denotes the complement of set X. Grey level extensions can be defined by the union of different area operations in consecutive thresholded sets [42].

- **Granulometry and Area Granulometry**

  Granulometry is a useful tool for obtaining *a priori* information about the size of the objects in the image. Granulometric size distributions are computed via a family of openings which have increasing, anti-extensive, idempotence properties [43, 44]. Since the area opening attribute (area) has the above-mentioned properties, the granulometry can be computed with successive area openings [45]:

$$G_\Lambda(f) = \sum \gamma^a_{\lambda_i}(f) - \sum \gamma^a_{\lambda_{i-1}}(f) \qquad \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\} \tag{1.11}$$

  where $\gamma^a_{\lambda_i}(f)$ is area opening of image f (binary or grey) with area threshold λ_i.

- **Top-hats**

  The arithmetic difference between an image and its opening or closing is referred as morphological top-hats. Top-hats by opening (white top-hat, WTH) and by closing (black top-hat, BTH) are defined as below [29]:

$$WTH(f) = f - \gamma_B(f) \tag{1.12}$$

$$BTH(f) = \phi_B(f) - f \tag{1.13}$$

  Since opening and closing operations are anti-extensive and extensive operations respectively, the white or black top-hat is always greater than or equal to zero. The white top-hat extracts the brighter content of the image which is removed by opening. On the contrary the black top-hat extracts the darker content which is removed by closing. The extension to white *area* top-hats and black *area* top-hats is performed by using area opening and closing operations respectively [42].

- **Reconstruction**

  The reconstruction operation can be easily understood after introducing the *markers*. Many MM operations are extended to operate by an additional set (image) which marks particular components in the input (i.e. geodesic transformations). Thus, instead of one input, with two input images operations can

provide selective or conditional behaviours for elementary operations. For example, a marker controlled dilation operation (i.e. geodesic dilation) is defined as follows [46]:

$$\delta_g(f) = \delta(f) \wedge g \tag{1.14}$$

where the elementary dilation operation is applied to the marker image (f) and then its point-wise minimum is calculated with the mask image (g). Therefore, the mask image limits the dilation operation. Since, they are limited by the mask image geodesic operations can be iterated to reach stability which leads to the definition of the morphological reconstruction. The reconstruction by dilation operation is defined by iteration of elementary dilations until stability [46]:

$$R_g^{\delta}(f) = \delta_g^{(i)}(f), \quad \text{where} \quad \delta_g^{(i)}(f) = \delta_g^{(i+1)}(f) \tag{1.15}$$

The reconstruction by erosion is defined in a similar basis [29, 46].

- **Regional Minima and Maxima**

  A *regional maximum* ($RMAX_h$) in a given value (h) of an image (f) is the connected components of pixels that have neighbour pixel values strictly lower than h. Thus, it is different than the *local maximum* that is defined by a pixel p that neighbours have values which are lower or equal to $f(p)$. The regional maximum of all h values, in an image is referred as regional maxima can be computed with reconstruction by dilation as follows [46]:

$$RMAX(f) = f - R_f^{\delta}(f - 1) \tag{1.16}$$

  In a similar way, the regional minimum ($RMIN_h$) in a given value (h) of an image (f) is the connected components of pixels that have neighbour pixel values strictly higher than h. The regional minima can be computed with reconstruction by erosion as follows [46]:

$$RMIN(f) = R_f^{\epsilon}(f + 1) - f \tag{1.17}$$

The further details of derivation and proofs can be found in [29,46].

- **Watershed Segmentation**

  The watershed transformation is a powerful morphological image segmentation tool. In a topographic representation of an image, the grey level value of each pixel can be perceived as an elevation at this point [47]. The watershed transformation, in an analogy to flooding simulations, tends to create partitions of the image corresponding to *catchment basins*. When the objects in the image correspond to minima, and the object boundaries are presented, the output of the transform divides the image plane into unique regions (influence zones of minima) associated with the objects.

  Here, we include Soille's watershed definition in terms of flooding simulations briefly, see [29] for a complete description.

  The watersheds of an input grey level image f correspond to the boundaries of catchment basins of f. Let $rmin$ denote a regional minimum of f and let show $CB(rmin)$ the catchment basin associated with this $rmin$. The points of $CB(rmin)$ which are on or below grey level h can be shown using its intersection with the threshold set at level h:

$$CB_h(rmin) = CB(rmin) \cap T_{t \leq h}(f) \tag{1.18}$$

  where $T_{t \leq h}(f)$ denotes threshold result of f at level h. The subset of all catchment basins which have grey scale value less than or equal to h can be shown as:

$$X_h = \bigcup_i CB_h(rmin_i) \tag{1.19}$$

  where the subset of regional minima at level h can be denoted by $RMIN_h$. Hence, at the lowest grey scale value $hmin$, the set of catchment basins is equal to the threshold set and also to the regional minima at this level, i.e.

$$X_{hmin} = T_{hmin}(f) = RMIN_{hmin}(f) \tag{1.20}$$

The catchment basins in the next level can be found using the union of the newly emerged minima and influence zones of the previous minima:

$$X_{hmin+1} = RMIN_{hmin+1}(f) \cup IZ_{T_{t<hmin+1}(f)}(X_{hmin}) \tag{1.21}$$

where $IZ_{T_{t<hmin+1}(f)}(X_{hmin})$ are influence zones of the catchment basin subset X_{hmin} in the grey level threshold set $T_{t<hmin+1}(f)$. Influence zone $IZ_A(K_i)$ of a connected component K_i in A is the connected set of points of A whose geodesic distance to K_i is smaller than their distance to any other component (see [29]).

The set of all catchment basins of f is equal to the set X_{hmax} where $hmax$ is the maximum grey level of f, and therefore can be calculated iteratively:

$$X_{hmin} = T_{hmin}(f), \tag{1.22}$$

$$\forall h \, [hmin, hmax - 1], \tag{1.23}$$

$$X_{h+1} = RMIN_{h+1}(f) \bigcup IZ_{T_{t<h+1}(f)}(X_h)$$

Finally, the watershed lines are defined as the boundaries of X_{hmax}.

1.9 Pattern Recognition

Pattern recognition which drives machine perception is a field in the broader paradigm named *Machine Learning*, which is concerned with learning of the computers. Machine perception has a wide range of application areas such as manufacturing, medical examination, authentication, surveillance, astronomy or military applications. It has active research areas such as object recognition, speech recognition, face recognition, optical character recognition, DNA sequence analysis and many more. Basically many of these recognition tasks only differ in the subject of study but are related by a similar set of mathematical notions which are pattern classification techniques. Below, some key definitions and functions are provided that are common among the most classification techniques. More information can be found in various sources, e.g. [30–33].

- **Feature**

  Suppose we want to automate a classification task which is performed manually. First thing we need to define is a set of observations or measurements which characterise and discriminate the entities (e.g. objects, faces, sound) that we want to classify. This set is called the *feature* and may be comprised of many individual *features*. A process or function that analyses the signal coming from the object and producing a set of features is a *feature extractor*. Suppose a classification of circular and square like objects is to be performed, and the object signal is a 2-d binary image, a feature extractor may produce the area (A) and perimeter (P) measurements as the features to be used in the *classifier*.

- **Classifier**

  A classifier is a function which assigns a given feature vector associated with the object to one of the classes that are defined in the problem. The simplest classifier can be a rule-based or heuristic-based function which performs the decision based on an if-then logic. If we use our last example, a classifier would be a statement such as: If $4\pi A/P^2$ ratio is 1 then assign object to circular objects class, otherwise assign object to square objects class. However, in practice the problems are more complicated for simple heuristic decisions. Instead most classifiers are designed to perform decisions based on more complex functions which are statistically justified on a set of observations (*training samples*).

- **Training Samples and Learning**

  More complex problems require bigger feature vectors comprised of more features which result in a higher dimensional *feature space*. Therefore classifier decisions require complicated evaluations. Usually the task of the classifier transforms into calculation of the conditional probability of a given feature vector for the classes.

  In general, learning is an operation to reduce the classification error with an algorithm based on a mathematical model which refines the decision function.

This is often performed using a set of pre-labelled (supervised) examples (*training samples*) which the decision function can be optimized on. The optimisation of the classifier decision function parameters with a labelled set of examples is often referred as *supervised learning*. In some problems it may be impossible to provide label information (supervision) for the training samples. A classifier to solve such a problem would perform a *clustering* on the training set. This process is called unsupervised learning or clustering.

If we have a metric feature space the decision function defines a decision boundary in feature space which is a partition for the classes. In simpler problems the decision boundary can be a linear hyperplane, however in practice most problems require non-linear and more complex decision surfaces. A simple decision function can be based on the histogram of the features (class conditional distributions) calculated for the each of the target classes. The classifier can use the histograms of the features (density estimation) to produce a likelihood ratio. If we have the *a priori* probabilities of the classes the decision can be made according to the Bayesian decision rule: Decide w_1 if $P(w_1|x) > P(w_2|x)$ where $P(w_j|x)$ denotes the posterior probability of the state of being w_j given that feature value x [33].

- **Generalisation**

  By learning, a classifier aims to improve a decision capability for previously unseen samples. This is known as *generalisation*. The learning process using training samples tends to reduce the classification error on training samples. However, training samples may not be optimal representatives to provide a good generalisation. Also a poor generalisation can occur if an iterative learning model *over-fits* to the training samples. Hence there is an approach named *validation* which can control the fitting.

- **Validation**

  While iteratively reducing the classification error on a training set, the valida-

tion is performed by evaluating the trained classifier on previously unseen set of samples (validation set). The training is stopped when the error on validation samples starts to increase. This prevents over-complex models from *over-fitting*.

- **Evaluation and Error Measures**

  Testing aims to evaluate and report the generalisation capability of the classifier for the problem. It can also provide a comparison of different factors such as discrimination capability of the features and may suggest possible improvements. An average classification error value on the testing set can represent the success of the classifier. However, in most problems more detailed error analysis is required for the analysis of the performance. In binary classification tasks the number of true classifications versus false classifications can be used to plot Receiver Operating Characteristics Curve (ROC) [33]. ROC analysis is a useful tool for comparing and visualising the classification performances [48, 49].

Chapter 2

Literature Review

In this chapter we will present the related literature on the basics of automated microscopy imaging, blood image analysis, and Computerised Diagnosis of Malaria (CDM). CDM is not a highly active research area due to the narrow focus on the disease which requires interdisciplinary research. However, there are a remarkable number of studies concentrated on processing of blood cell images, mainly aiming at White Blood Cell (WBC) analysis. Some of these studies which investigate the use of different features and classifiers are included in this review.

2.1 Processing of Blood Cell Images

2.1.1 Imaging

The importance of digital imaging has grown in medical applications recently. In addition to computer aided diagnosis, it enables various applications such as remote intervention [50], telemedicine/telepathology [51,52] or education [53].

A full automated CDM system requires complete digital control of the microscope while remote control can have applications in a semi-automated system. In [54] the authors report a successful method to control a microscope using a digital control board and 3 stepper motors to enable remote observing of an expert microscopist of a slide prepared by a technician. However, even when full control of the microscope is provided, the required time for capturing a whole slide could be too much to be feasible. The number of images required to completely capture a blood slide is a function of field of view of the imaging system. In [55] the required number of images to capture a $2cm^2$ area specimen at 20x magnification is calculated to be nearly 1300 images using a 1300x1030 pixel 2/3 inch CCD camera. CDM requires 100x objective magnification, so the number of captured images would be twenty five times higher.

Hence, it corresponds to over 30000 slide movements, focus and CCD sensor shutter operations. To reduce time requirements they [55] propose to capture the images while the slide is continuously moving which introduces another problem of image blurring. They propose to use Xenon strobe lights instead of conventional lights to solve this problem, which raises the cost of the system dramatically.

Finding a fast image capture technique for a low cost configuration system remains unsolved, but probably a human expert will require more time to manipulate a slide and focus the microscope to observe 30000 of fields. Hence, the number of fields the expert would examine is smaller. In WHO malaria microscopy tutorial [9], examination of only 100 fields is recommended before giving a negative decision. Additionally, if a parasite is observed in a field, 100 more fields (or 200 WBCs, 0.025μl of blood) would be sufficient to calculate the parasitemia in thick films. Since it is time consuming and the sensitivity threshold is high, routine examination of thin blood films is not recommended for the positive/negative type of diagnosis. However, if parasites are found, examination of 25 fields (average 200 RBC per field yields 5000 red blood cells) would be sufficient to calculate the parasitemia in thin blood films.

In CDM the observed fields can be limited with 200 fields or alternatively with the number of fields to capture a fixed number of Red Blood Cells (RBCs) (WBCs in thick films). For example, with a 100x objective, capturing the images of 200 fields with an average 200 RBCs (depends on the CCD size and the total magnification between specimen and CCD) may collect 40000 RBCs which corresponds to approximately 0.008μl of blood in a thin film. This could provide a sufficient number of cells for the sensitivity threshold of the expert examination (0.004μl) (manual microscopy threshold: in Section 1.3.1). However, the blood smear is not homogeneously distributed on the thin blood film which suggests that a systematic approach would be necessary for field selection. In [26] a method was proposed to enable automatic detection of the appropriate fields for the post-processing. The estimation of "appropriateness" was established following the criteria of a good field in the practice of manual microscopy: RBCs are

just contacting. This was achieved by an algorithm composed of MM operations such as granulometry and distance transform. In this study, we propose a method which can process images regardless of the field density, and thus removes the necessity of employing a selection procedure (see Chapter 6).

2.1.2 Colour Constancy

It is very common for automated classification methods to employ an initial stage to reduce the effect of some expected variations. Various properties of the microscope imaging setup can cause variations in observed colour, intensity of the image or size of the cells. These variations that can be easily tolerated by an expert microscopist may not be tolerated by a computer algorithm. If there is only one imaging setup or if the system can be calibrated for colour variations, this would be a minor concern. However, images acquired from different setups may have negative effects on the accuracy of a computer classification.

Yagi *et al.* [56] discuss the issue of image standardisation for pathology in telemedicine. According to the authors, a medical imaging standard can enable the systems to share the image files. However, the current obstacles are human and imaging factors. The human factor is related to the differences in preparation of the specimens (i.e. staining) while imaging factors are related to the sensor characteristics, lighting or different calibration.

There are various studies addressing the calibration and colour constancy issue in imaging for general machine vision purposes [57,58]. However, it is difficult to say the same for microscope imaging which requires special approaches.

In machine vision, an intuitive method for colour calibration is to use a reference colour chart and then tune the system according to the differences in the observed colours. There are some colour calibration methods in the literature which are specifically proposed for medical applications. In [59] authors proposed a method for colour

normalisation for tongue images comparing the response to a reference colour chart. The calibration was performed using a support vector regression classifier [33] which assigns the colours to appropriate bins on the colour reference chart. Similar reference colour chart comparison methodologies were followed for calibrating skin images for dermatology imaging in [60, 61]. A more comprehensive method was proposed in [62] a method for calibration again with response to a reference colour chart. They have generalised the image formation model based on Lambertian surface assumptions and estimated the illumination spectrum while spectral surface reflectivities were determined by Wiener inverse estimation [63].

However, there is an important condition of microscope imaging which impedes calculations based on the Lambertian surface model and use of the reference colour charts. The surface reflectance models are not appropriate because the sensor (or human eye) does not receive light reflecting from a surface. The light reaching the sensor is the attenuated light which is left from object's absorption. In fact, image formation of the slides under light microscopes more appropriately should be modelled with the "Beer-Lambert Law" which states that there is a linear relationship between the concentration and thickness of illuminated media and the "absorbance" [63]. Additionally, the reference colour patches which have to be in a transparent form with a perfectly homogeneous colour are not practical for microscopes. Even it was possible to manufacture them; there is still the previously mentioned human factor in preparation of the slides which results in non-standard and inhomogeneous staining concentrations.

The problem of non-standard preparation of the slides was addressed in [64,65]. To correct under/over staining conditions of the slide, they obtained the spectral transmittance by a multispectral camera: a camera equipped with different band filters to capture the spectral reflectance on separate bands. They mathematically modelled the relation between the transmittance and the amount of stain (dye) for each pixel using the Beer-Lambert Law and Wiener inverse estimation [63]. If we overlook the cost of a multispectral camera, it [64] is an important study providing a mathematical model

of the staining concentration-transmittance relation which enables digital correction of non-ideal stain concentrations. However, the variations due to the different camera parameters and light sources were not addressed which leaves the camera side of the problem unresolved. Nevertheless, for the CDM application, we may not have the luxury of adding the cost of a multispectral camera to the system; it is not practical to capture many (e.g. 10) different bands of the same field to estimate the amount of dye.

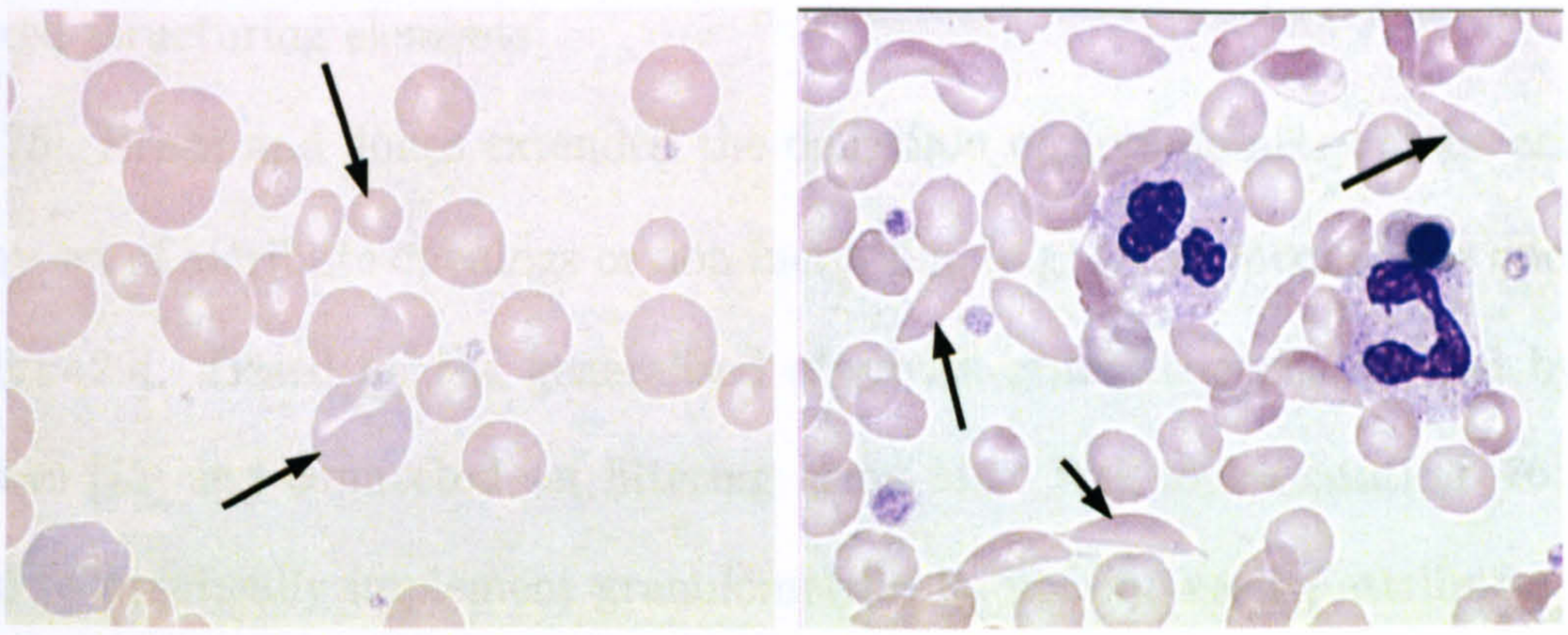
In this thesis, we propose a very practical method ensuring the colour constancy on the images captured from various (unknown) sources. The method exploits the special characteristics of the peripheral blood images. It is explained in detail in Chapter 5.

A final issue that could be appropriate to mention here is the choice of colour space. Most imaging devices produce a colour image in RGB colour space. Many studies such as [34,66,67] suggest a transformation into alternative colour spaces such as LUV, CIE, HSV in which they claim easier interpretation or lower correlation between separate channels [68]. However, there is no evaluated study that demonstrates one colour space is more favourable to the other for the purpose of any particular blood image processing problem.

2.2 Size of the Cells

Average sizes of healthy human peripheral blood components have known limits. Normally, the diameter of a red blood cell and platelet is between 6-8μm, 2-3μm, respectively. WBC size can vary between 8-20μm depending on the type. However, there is a normal tolerance of the size deviation on the averages values. For example, normal size deviation for RBC is between 10-15% in healthy blood [69].

If known, the CCD pixel resolution and the magnification (between specimen and CCD) can be used to calculate expected sizes of the blood cells that are present in the image. Additionally, in normal healthy blood RBCs can be expected to have a regular circular shape. However, there are some conditions (i.e. anaemia) which result



(a) (b)

Figure 2.1: Examples of abnormal cell size and shape (a) anisocytosis (b) sickle cell [71]

in abnormal shapes and sizes. For example, in anisocytosis RBC size can greatly vary and in sickle cell disease the shape of the red blood cells are deformed. Figure 2.1(a) and 2.1(b) show examples of anisocytosis and sickle cell, respectively. Similarly, there are various disorders affecting other blood components which result in abnormal shapes and sizes [70].

Therefore a prior analysis may be essential to estimate the size or the shapes of the cells. As mentioned earlier in Section 1.8, mathematical morphology provides useful functions which can be used in object level processing of images. One of the most useful functions is granulometry (pattern spectrum) which can provide the size distribution of the image. Granulometric size distributions are computed via a family of openings which have increasing, anti-extensive, idempotence properties [43, 44]. Though the definition of granulometry does not suggest any special type of opening operation, in practice it is usually implemented via a set of increasing width structuring elements of a pattern (e.g. square, disk, and hexagon).

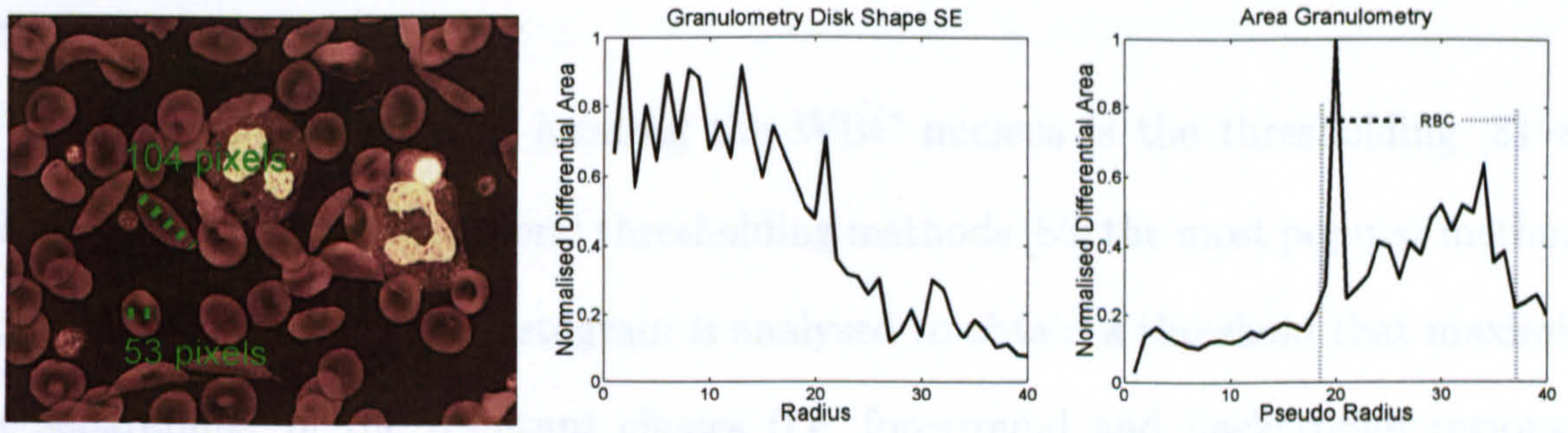
Granulometry with circular structuring elements has been employed in some studies as a pre-processing stage to estimate the average size of the cells [26, 34, 72–74]. However, the abnormal conditions as shown in Figure 2.1 and the holes inside the cells degrade the accuracy of a granulometric size distribution calculation when calculated

with fixed structuring elements.

In [75], Breen and Jones extended the definition of granulometry to be calculated with any set of attribute openings or non-increasing (e.g. area) opening like operations: thinnings [29]. Based on the generalised attribute granulometries defined by Breen and Jones [75] and connected set filtering using Max-Tree representation [76], it was possible to practically implement granulometries by non-increasing attributes. In [77] the authors proposed an implementation of shape pattern spectrum which was later extended to the calculation of 2-d granulometries (shape X size) in [78] and to the vector granulometries in [79].

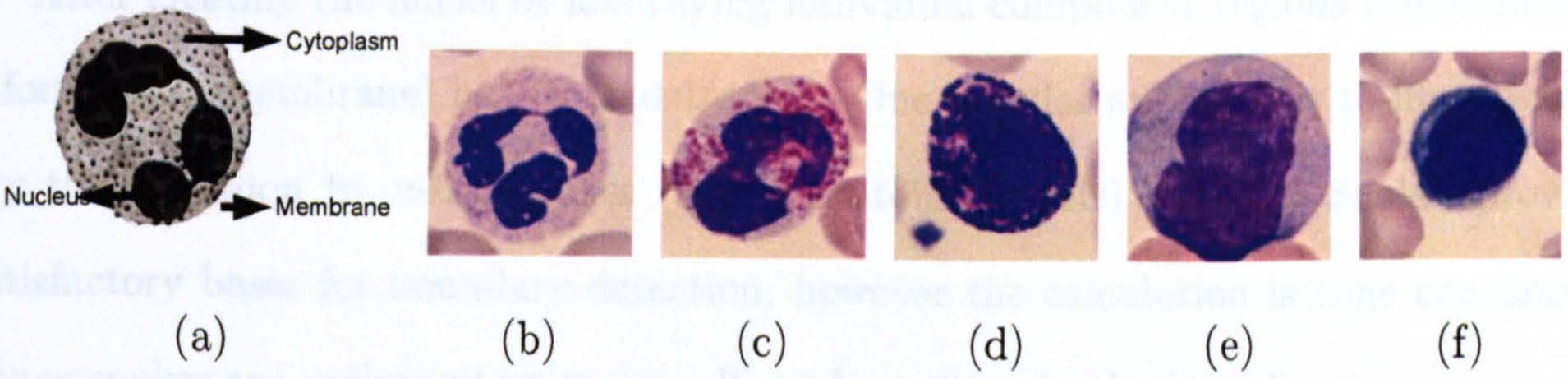
To cover deformed non-circular shapes of cells, Rao *et al.* [45] proposed the utilisation of area granulometry instead of granulometry calculation with fixed structuring elements. Area granulometry was then used in different studies as an improved size estimator for the blood cell images [35,39,40]. In these studies, area granulometry was calculated with a series of openings which is not efficient. A method for fast computation was proposed earlier in [80] based on a connected sets tree (i.e. Max-Tree [76]) representation. Figure 2.2 shows the plots of granulometry and area granulometry calculations on the grey level negative of the sickle cell image that is shown in Figure 2.1. In this image, RBC diameter changes between 20-35 pixels. Since RBCs are not all circular and homogeneous (have holes), granulometry is confused; area granulometry is more accurate.

In this thesis, area granulometry is used to obtain size distribution for the suggested methods of segmentation and colour normalisation (Chapter 4 and 5, respectively). Moreover, in the classification experiments, the local area granulometry is employed as a shape descriptor feature for stained object classification (i.e. parasite detection) (Section 7.2.3).



(a) (b) (c)

Figure 2.2: Granulometry vs. Area Granulometry: (a) Negative of the grey level sickle cell image, (b) granulometry, (c) area granulometry



(a) (b) (c) (d) (e) (f)

Figure 2.3: White blood cell components and types: (a) an illustration of WBC showing individual components. Examples of 5 main WBC types: (b) Neutrophil, (c) Eosinophil, (d) Basophil, (e) Monocyte, and (f) Lymphocyte

2.3 Segmentation

Segmentation is the common basic task in blood cell analysis applications. There are a huge number of references to blood cell segmentation in the literature. Due to its popularity in haematology applications, most of these only refer to WBC segmentation and related tasks (e.g. counting, classification). Since the WBCs are present in the peripheral blood images and they are also highlighted like malaria parasites with the Giemsa stain, the methods that are developed to segment, identify, and classify them are important for this survey.

WBC consists of three main structural components: nucleus, cytoplasm, and cell membrane (Figure 2.3(a)). WBC segmentation most commonly refers to first locating the WBCs (on the stained image) then segmenting to locate their components for

further analysis.

A common approach for locating the WBC nucleus is the thresholding [81–84]. Though there are various general thresholding methods [85] the most popular method is Otsu's [86] in which image histogram is analysed to obtain a threshold that maximizes the separability of the resultant classes (i.e. foreground and background regions or stained and non-stained pixels). An alternative method to thresholding is to use a k-means clustering [31] or mean shift clustering to obtain separate regions representing the different structural components of the WBCs and other components [66, 67].

After locating the nuclei or identifying individual component regions the boundary information (membrane) has to be extracted. One popular approach is to fit a boundary to the region by using snakes (active contour models) [87–91]. Snakes provide satisfactory bases for boundary detection; however the calculation is time consuming (since snakes are optimised on every cell) and sensitive to the initialisation.

For the CDM and similar applications, the important segmentation task is to find the unique regions associated with individual blood cells (RBC, WBC, and platelet). In other terms, it corresponds to locating each cell in the image independent of their type. In [82], the authors propose a method that is using trainable Gaussian boundary models for cell image segmentation which they have tested on the images taken for cancer screening. However, calculating each cell boundary, fitting a Gaussian model and matching can take a significant amount of time.

On the contrary, again mathematical morphology provides very practical tools for segmentation of blob like objects. One simple method is to use watershed transformation which was described in Section 1.8. In watershed segmentation the image is seen as a topographic surface. In an analogy to the flooding simulations, watershed transform calculates unique regions associated with each regional minimum (catchment basins) which results in excessively over-segmented objects. In [92] the authors proposed estimating the quality of watershed segmentation on the object based on the calculation of a Mahalanobis distance to a cell model which then used to exclude under

qualified segmented objects from further processing. In [93] the authors have proposed to merge of initially found catchment basins based on their depths and areas to find actual object regions.

For a simple blob image segmentation problem, watershed transformation can be used together with binary image distance transform to mark and segment only the important minima (corresponding to objects) [47]. Yet, the distance transform do not solve the problem completely. In [94] the authors proposed to filter the distance transform to improve the watershed segmentation. They have evaluated the counting performance of the system, however, the results were not correlated with manual counting. In [72], Di Ruberto *et al.* proposed a method assuming the RBCs are circular. They suggested that overlapping can be separated by a structuring element of fixed size which was obtained from the granulometry. As we have seen in the previous section, the circularity assumption always does not hold. However, they have addressed the presence of the white blood cells which was affecting granulometry. They have detected WBCs and parasites using a threshold on the image histogram and excluded prior to granulometry calculation to improve the size estimation for RBCs.

Later, in several publications [35, 95] Rao *et al.* addressed the over-segmentation under-segmentation problem. In [95], they proposed an iterative segmentation using the distance transform in which the under-segmented cells were detected by checking their area ratio to the estimated average cell area. Average cell area was estimated by area granulometry (i.e. peak index). In [35], they proposed to use area top hats for the marker extraction instead of distance transform. However, the presence of WBCs (which are larger than RBCs) were neglected that resulted in over-segmented WBC regions.

In this thesis, to overcome the over-segmentation in the watershed transform a modification is proposed. By this modification a new MM tool is introduced: attribute watershed transform in an analogy to the attribute openings and thinnings [75]. A novel cell image segmentation method is proposed by employing this attribute watershed

transform incorporating the Radon transform [96] to extract markers. The cell counting performance was evaluated by comparing to manually prepared ground-truth data. The presence of WBCs was taken into account and they were marked to prevent over-segmentation. The details of the method can be found in Chapter 4.

2.4 WBC classification

As mentioned in previous section, WBC analysis is a popular research subject. WBCs are one of the components affected by Giemsa staining and they are in the non-parasite object class for CDM. Hence, the methods developed for automatic classification of WBCs are highly relevant to CDM and to this survey. The main contribution of reviewing WBC classification methods is to deduce the motivation of choosing the features and classifying methodologies which can be incorporated into the CDM. Additionally, some studies that are reviewed below aim the classification of RBC, WBC, and platelet into separate sets which are also important for CDM.

WBCs (leukocytes) are categorised into 5 main types: Neutrophil, Eosinophil, Basophil (polymorphonuclear); Monocyte, Lymphocyte (mononuclear). Figure 2.3 shows an example for each category of leukocyte. The different types of WBCS are distinguishable by their nucleus shapes, sizes, cytoplasm or presence of granular stippling when observed on a Giemsa stained thin blood film slide [70]. It is possible but harder to distinguish them on the thick film slides due to the de-hemoglobinisation of the cytoplasm during staining process.

In [97] the authors propose using cell size (number of pixels inside the cell boundary), and average value of the red, green, blue colour channels to classify RBCs, WBCs, and platelets into separate classes. They have trained a back-propagation neural network (BPNN) [32] using these features that are extracted from 100 cells and report 0% error on the training set. However, neural networks have a well known property of converging to any decision function provided they have enough number of neurons

and layers [33]. Hence, the 0% training error without any evaluation on a separate set does not suggest any generalisation.

Aiming at the same classification, in [98], the features were extracted by calculating a discrete wavelet transform, then training with a BPNN and ALOPEX [99] algorithm separately and performances were compared. The recognition rate on a test set which was a portion of 11 images was reported to be 49% and 89%, respectively. The cells were manually labelled and pre-processed. However, the wavelet transform provides direction sensitive features (rotation variant) which may demonstrate poor performance in cell classification.

In [83] a classification accuracy of 91% was reported on a set of 71 WBCs. The classification was done based on the mean values of 10 features (no reference was given to a classifier description). Features included measurements such as nucleus area, cytoplasm area, average colour, compactness etc.

In [100], the authors proposed eigencells in an analogy to the eigenfaces [101] which is known to be size and rotation sensitive. They suggested that the *direction* of the cell can be found on the line connecting the cytoplasm and nucleus centre of masses. However, this assumption may not hold when the WBC does not exhibit clear regions of cytoplasm and nucleus. Even if a way can be found for calculating direction for the WBCs it is not applicable for *Plasmodium*, platelets or artefacts.

Hence, there exist more popular features which are rotation or scale invariant. In [102] the authors presented a WBC image indexing and retrieval system (outputs the closest match from a database) using a K-Nearest Neighbour (KNN) classifier on the several features extracted from the nucleus. The features were comprised of Fourier boundary coefficients, nucleus area and a texture feature. The system was not intended to perform automatic (WBC disorder) diagnosis but provide decision support for the pathologist.

In [103] a system was proposed to perform the automated differential count of WBCs. They have calculated a 57 element feature vector including several measure-

ments on the colour, shape, texture of the cytoplasm and nucleus. The feature set included affine invariant moments, colour co-occurrence histogram, colour histogram, area, perimeter, compactness etc. They implemented three different classifiers: KNN, learning vector quantization (LVQ), BPNN and support vector machine (SVM). On an image set containing 258 WBCs (train-test 7/3), they have reported test set accuracies of 81%, 83%, 90%, 91% respectively for the KNN, LVQ, BPNN, and SVM classifiers.

In an earlier study [104], shape features of cytoplasm (area, perimeter etc.) were found to be not discriminative (and excluded): only the chromatic features (colour based) were extracted. However, for the nucleus they have used both types of features. The chromatic features included statistical measurements such as standard deviation of the nucleus and cytoplasm of the values in colour channels or the ratio of channel averages; the morphological features included measurements such as nucleus area and compactness. Several BPNNs were trained with the feature vector of 7 elements; the overall class decision was given by majority of votes. They have reported 98% accuracy on a test set including 662 WBCs obtained by two different imaging systems.

In [105] a set of 23 features (similar to the previously mentioned ones) were used to train KNN, BPNN, and parallel BPNN classifiers. A mean classification error was reported to be 0.08% with the combination 5 parallel BPNN classifiers where errors for the KNN and single BPNN were 0.14% and 0.18%, respectively. However, they have used a total of 134 WBCs and the accuracy of the method was measured with 10-fold cross validation: the training was performed on 9/10 of total images by validating on the rest of the set.

In [106], besides area and circularity, some innovative features such as curvature, skeletons and multiscale fractal dimension were proposed for automatic leukaemia diagnosis (WBC disorder, i.e. type of blood cancer). However, they have only analysed the feature relevancies checking the scatter plots on set of 75 healthy WBCs; there were no evaluations for classification to suggest the performance of the new features (e.g. skeletons, fractal dim.).

In [107] different classifiers were compared. A set of 25 features composed of five nuclear area and perimeter measurements and 20 colour based features measuring minimum, maximum, and mean of RGB channels and their ratios (e.g. mean R/ mean G) were extracted. Using 200 WBC images, 7 different classifiers were trained including BPNN and Naive Bayes. The best recognition performances were obtained by BPNN (1.95%) and Naive Bayes (4.0%) on the tests of 206 WBC images.

In [108], a total set of 62 individual features were used to train a SVM classifier to classify 5 different types of leukocytes and the malignant type of lymphocytic leukemic cells. They compared the classification performances of SVM to the Naive Bayes and concluded the former was superior. The dataset included over 900 images: 600 were used for training and 300 were used for testing.

In order to finalise the review on WBC classification the feature choices and classifier performances can be summarised. The most common features can be divided in two categories: morphologic (shape) and chromatic (colour). A common choice for morphological (shape) features is the area and perimeter of the nucleus and derived measurements based on their ratios such as compactness. On the contrary there is no common choice for colour based features; however they are extracted from histograms or statistical measurements of the separate channels. However, there is no study reviewed here comparing individual feature performances that can motivate selection of a subset of features from the overall features used among the different experiments.

On the classifier side, it is again difficult to suggest one will provide more generalisation than the other due to the fact that separate datasets were used in evaluation. Though some studies compare the classifier performances, different classifiers can be tuned with their parameters to achieve better results. Or some simple classifiers such as KNN are highly dependent on the choice of distance metric: there can be significant performance difference between a KNN using Euclidean distance and another using city block distance metric (see Section 7.4.1.3). As a matter of fact, a theorem of general pattern recognition suggests that there is no superior classifier [33]. However,

performance of a classifier in a specific task can be evaluated: by comparing to the classification of an expert (manual); or the compared classifiers should be trained (i.e. estimate own parameters) automatically.

Interpreting the classification results on their own, it would be too optimistic to conclude automated WBC classification on the stained images can be done with accuracies over 90%. Most of the studies presented here do not take into account immature cells, platelets or artefacts. Some of the results were produced using small sample sets which may be over simple compared to an actual diagnosis scenario. Moreover, the problem of colour constancy was rarely addressed, which may hinder practical implementations.

2.5 Diagnosis of Malaria

In this section we will review the studies aiming at CDM. The topic has been subject to other research studies in the University of Westminster. First, Dr. Cecilia Di Ruberto, later Dr. K.N.R. Mohano Rao has studied on automated thin film examination with Dr. Andrew Dempster. Previously under the size estimation (Section 2.2) or the segmentation sections (Section 2.3), some of their contributions were reviewed. Apart from these auxiliary tools they also investigated parasite analysis.

In [73], Di Ruberto *et al.* proposed an automatic thresholding method to extract malaria parasites based on granulometry and regional extrema. The method converted RGB images to HSV (Hue, Saturation, Value) colour space, then employed morphological operations on hue and saturation channels. They suggested that the markers for the parasites together with the WBCs can be detected by calculating the intersection of the regional maxima obtained from the two separate channels. Then they refined the detection process by using a further thresholding using the mean values of pre-detected pixels as thresholds. A feature extraction or classifying strategy was not proposed that could discriminate between parasites and WBCs. Moreover the platelets and artefacts were not considered. This method had two central assumptions that hinder its use as

an algorithm for malaria parasite detection in peripheral blood images:

1. *The regional maxima operator detects parasites and WBCs*: The regional maxima operator finds the pixels on the image which have all neighbour pixel values lower or equal to its value. Even when the image has no stained pixels or parasites, regional maxima will clearly mark some pixels according to its definition.

2. *Stained pixels belong either to the parasites or WBCs*: All the stained (highlighted-saturated) pixels stemming from different components: parasites, WBCs, platelets, artefacts will exhibit the same nature in the saturation image.

Later in [34], Di Ruberto *et al.* published a more comprehensive report on the subject. They used the same procedure as above to detect stained pixels. Then they proposed two different classification methods to analyse life-stages. Again platelets and artefacts were not considered, nor were the different properties of the different species. However, they eliminated the WBCs by applying a disk shape opening bigger than the average cell size and also detected and isolated schizonts by checking separated stained regions that are close to each other. For the first classification method, they proposed the use of the number of morphological skeleton end points as the features. The life stages were correctly classified in one image using the relation of circularity (estimated by the number of endpoints) of the cell and parasite maturity: more endpoints indicates non-circular objects; immature trophozoites (ring shapes) are circular; mature trophozoites are irregular; the gametocytes are more irregular. In the second classification approach, they have used cell region histograms as features and modelled the immature trophozoite and the mature trophozoite by calculating the histograms of 1 parasite for each class. They compared the counting performance of the system to the expert biologist markings and reported that the classification results on 12 images were more close to the both of the experts than the results of the experts were to each other. However, the evaluations were performed on the Plasmodium species which infects only rodents. In addition to the assumptions in their previous report, there were

two more assumptions which we can discuss here:

1. *WBCs can be detected by their size; schizonts can be detected by checking stained regions close to each other*: It is plausible to use the fact that WBCs are larger than normal red blood cells; however it is not safe to remove every stained object that is bigger than average RBC size. For example, in *P. Vivax* infection a mature schizont can be enlarged up to 2.5 times the size of a healthy RBC. Secondly, schizonts are not artefacts that should be removed from the analysis. Moreover, removing the stained objects that are close to each other would probably remove many other parasites. For example, in *P. Falciparum* infections it is very common that two separate parasites occupy the same cell.

2. *Immature parasites are circular, mature parasites are more irregular*: The illustrations presented in Figures 1.1- 1.4 show that this assumption will not always hold. If a categorisation could be based on the circularity of the infected cell, it would be appropriate to state: all life stages of *P. Malariae* and *P. Falciparum* (except the gametocyte stage) are more circular than *P. Vivax* and *P. Ovale* life-stages.

In his doctorate thesis [35], K.N.R. Mohano Rao addressed various issues of CDM. The stained pixel extraction method was improved by calculating the image bias and subtracting from the image before marking the stained pixels. This prevented misdetections when the images do not contain stained objects. However, though Di Ruberto *et al.* addressed the presence of WBCs and proposed an area threshold to remove them, Rao did not consider the presence of other stained objects: all the stained objects were assumed to be parasites. Therefore, all the stained objects were passed to the analysis of life-stage without checking if they are parasites. This assumption may hold on the *in vitro* samples (blood culture) but not for peripheral blood smears which are for diagnosis. The parasites (stained objects) were analysed to extract the nuclei count, nucleus area, parasite pigment to identify the life-stages. A rule based

classification scheme was developed for the *P. Falciparum in vitro* (not the peripheral blood) samples. The rules were based on area, the presence of the malarial pigment (Hemozoine) and nuclei count criteria [35, 109]:

- *merozoite*: If the radius of the circle of the parasite shape is less than 10% of the radius of the circle of the average cell and no pigment.
- *immature trophozoite*: If the area of the parasite is less than 70% of cell area and pigmentation does not exist.
- *mature trophozoite*: If the area of the parasite is less than 70% of cell area and pigmentation exists.
- *schizont*: If the area of the parasite is more than 70% of cell area and pigmentation exists. Mature-immature differentiation was made by nuclei count: if it was below 8 it was labelled as early schizont; between 8-16 as mid-stage schizont; above 16 as mature schizont.
- *gametocyte*: was not implemented.

The pigment presence and the nuclei count criteria follow the manual microscopy diagnosis practice of life-stage identification. However, the area threshold value 70% is difficult to generalise. Because it does not consider the major characteristic of *P. Falciparum*: more than one parasite can be present in the same RBC. For example, the area of two trophozoites can exceed 70%. The rules were only developed for the *P. Falciparum*. Other species were not studied.

The primary aim of CDM is to identify if there is any *Plasmodium* in the patients blood sample. A healthy person's peripheral blood has around 8000 WBCs and 150,000 to 450,000 platelets per μl blood. Any method which does not address the presence of other components (WBC, platelet) in the blood will clearly diagnose every patient as positive if these are identified as parasites. Furthermore, even if it is possible to use some heuristics for WBC and platelet elimination, artefacts are still a problem. It

is difficult to define an "artefact". Basically any stained object that is not associated with regular blood components or a disease can be named as an artefact: this includes RBC anomalies (e.g. Howell-Jolly bodies, iron deficiency), bacteria, spores, crystallised stain chemicals, and particles due to dirt.

However, a very recent publication [74] has addressed the problem. Ross *et al.* have proposed a two stage method for detecting the parasites. First, they extract all the stained pixels by thresholding (Otsu method [86]) by obtaining threshold values from the green channel histogram. They applied morphological operations similar to the ones in [73] (Di Ruberto *et al.*) to identify the objects that are stained. In the classification stage they proposed to use two different sets of features for parasite detection and species recognition. Using principal component analysis they reduced the set of features to 37, 38 which were including 75, 117 individual features respectively for parasite detection and species recognition. They have trained a two level BPNN for parasite detection and species recognition consecutively. The results for parasite detection were as follows: sensitivity (SE) 85.1% with a positive prediction value (PPV) of 80.8%. The specificity value or false detection rate was not reported. See Section 7.5 for descriptions of these measures, note that PPV is a population dependent performance measure. For the species recognition task the SE-PPV results were reported as follows: *P. Falciparum* 57%-81%, *P. Vivax* 64%-54%, *P. Ovale* 85%-56%, *P. Malariae* 29%-28%. The life-stage recognition problem was not investigated.

A more recent study [110] (Jan 2007) proposed "parasite counting" software. Similar to Rao's study [35, 109], they have experimented only on P. Falciparum *in vitro* images where the parasites are grown in a controlled environment (i.e. culture). Other regular blood components and artefacts, which are the main problems of diagnosis, are not present in *in vitro* samples. Thus, the proposed program counted the stained objects and compared the results to manual counting. They stated that the program was intended for research purposes; however, not applicable to diagnosis. The species and life-stages recognition problems were not studied.

Chapter 3

Data Collection

Unlike many visual pattern recognition studies concerned with ordinary subjects, the study of Computerised Diagnosis of Malaria (CDM) requires inter-disciplinary research. Visual recognition of familiar objects (e.g. face, car, and scene) can be studied by computer vision techniques without expert knowledge or special equipment. Some subjects such as medical image segmentation, research can also be performed without special knowledge. However, in CDM, the imaging study must be supported by medical expertise. There are two possible ways to research such problems.

The first obvious option is to cooperate with experts who have access to equipment; this was done in the early studies on the topic at University of Westminster (UoW). The images of thin blood films that contain parasites were externally supplied by National Institute for Medical Research (NIMR), UK. The research was not concerned with the pre-imaging setup such as blood slides or imaging conditions. The task was to develop algorithms which can perform an automatic analysis on the supplied images. However, the image collection was not sufficient to perform research on all the sub-problems in CDM. For example, as explained in Section 2.5, the presence of other stained objects was not considered in the earlier, since there was no additional data accompanying the images to locate/mark the parasites. Moreover, the images mainly contained parasite samples which infect only rodents and some *P. Falciparum in vitro* samples that are produced in a controlled environment.

The second option is to access the materials, equipment and knowledge directly. The materials are the blood samples and staining chemicals, the equipment includes a microscope and a digital camera, and the knowledge is the ability to identify the parasites using a conventional light microscope.

In the early stages of this research, the aim was to produce a low cost prototype CDM system incorporating a computer and microscope. Subsequently, a conventional

light microscope and camera were purchased for the project. In the mid-stages of this PhD program, the author attended a course on malaria which was intended and prepared for routine blood specimen examiners. The short course *Update on Blood Parasites* was presented by the University of Westminster (UoW), School of Biosciences (Jan-Feb 2005 [111]). This course taught staining methods for thin and thick blood films; thin and thick blood film screening for parasite detection; identification of species; calculation of parasitemia and identification of other blood parasites such as Microfilaria or Trypanosoma. Slides prepared during the course were then available for later use. After 8 weeks of laboratory practice, the author's diagnostic ability was assessed by the instructors along with other attendees and was found to be acceptable, at the average standard for the class.

Hence, slides used in this study that contain thin and thick blood films were mainly prepared as part of that course. Images of slides (apart from the ones supplied by NIMR) were acquired using the UoW microscope-computer setup (Figure 3.2(c)). The parasites and other structures in the images were manually identified and marked by the author but the parasite species that are present in the individual slides were confirmed during the training course. In the following sections the steps in this data collection procedure are explained.

3.1 Image Acquisition Setup

The overall scheme of the acquisition system is shown in Figure 3.1. It comprised a microscope, a digital camera and a computer. The computer is linked with the camera which enables remote viewing and capturing of the slide fields. The link is provided by a USB interface and TWAIN [112] software drivers. TWAIN is a popular camera driver interface which is supported by the usual operating systems and implemented in most of the consumer type cameras (e.g. web cam, digital compact). In the current system there are no motors on the microscope and no automatic control is possible

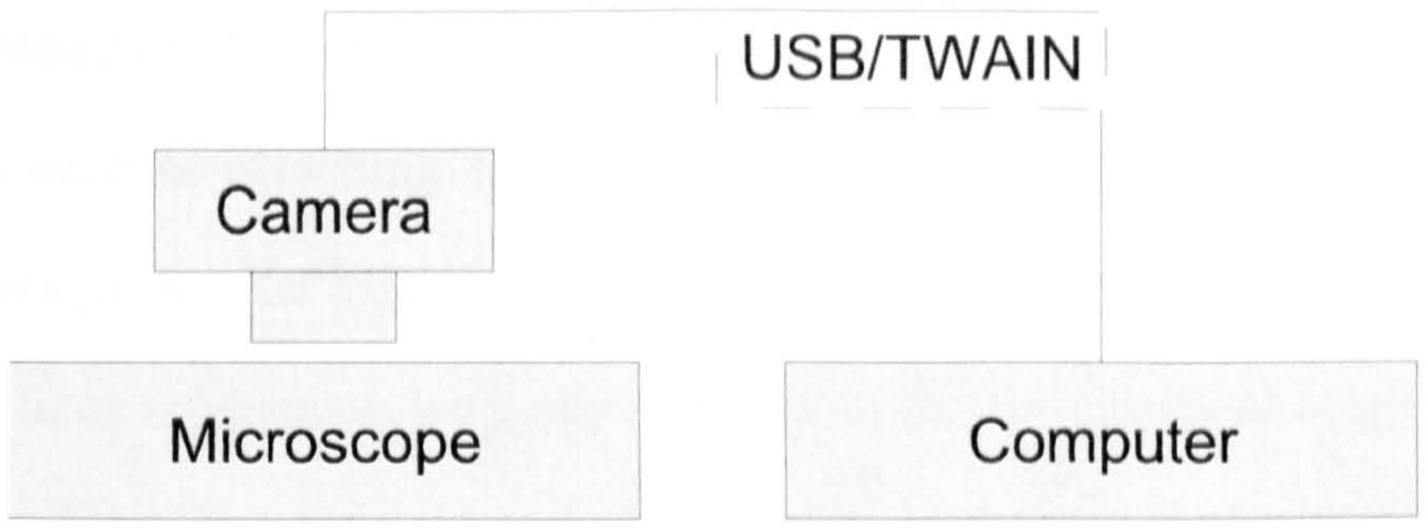


Figure 3.1: Imaging system components

for the slide positioning table or focus. Hence, there is no control link between the microscope and the computer.

The microscope used in this study is a *Brunel SP200 Research Microscope* (Brunel, UK). It is a low cost microscope (under £1000) but provides sufficient magnification (with 100x objective and 10x eyepieces: up to 1000x) and illumination (20 Watts) for the screening of malaria slides. It is equipped with a full field iris (aperture) which enables full control for the *Kohler illumination* setup. Kohler illumination is an important standard to set up microscopes to provide uniform illumination [113]. The importance of uniform illumination is explained in more detail in the following subsection. The microscope is shown in Figure 3.2(a).

Attached to this microscope, a mini camera is fitted into the microscope eyepiece without any additional equipment. The camera is equipped with a 1/3 inch CMOS sensor and can capture colour images of 583 x 438 pixels. It has a USB link and provides TWAIN compliant images. It is a low cost camera (under £100) and very practical for use with the microscope since it can fit directly into the microscope eyepiece hole. However, as will be seen later on some examples, due to the low quality of the image sensor, the acquired images were not suitable for CDM. Figure 3.2(b) shows the mini camera when it is attached to the one of the eyepieces.

Apart from this mini camera, we have also obtained some simple adapters which enable us to attach a consumer digital compact camera *Canon A60* (at a total cost of 100) to the microscope. The adapters consist of an optical coupler *Unilink* (Brunel,

UK) and a standard *Canon LA-DC52C* lens adapter for the camera which extends the lens and enables attaching to the Unilink optical coupler. Canon provided the *Software Development Kit* (SDK) of the camera which permits full control of the camera functions for later integration with any software in the development stage. The Canon camera is also TWAIN compliant and is supported by remote capture software. Figure 3.2(c) shows the Canon A60 camera with the Unilink adapter used to attach it to the camera tube of the microscope. This configuration can also be used in the microscope eyepiece as the mini camera.

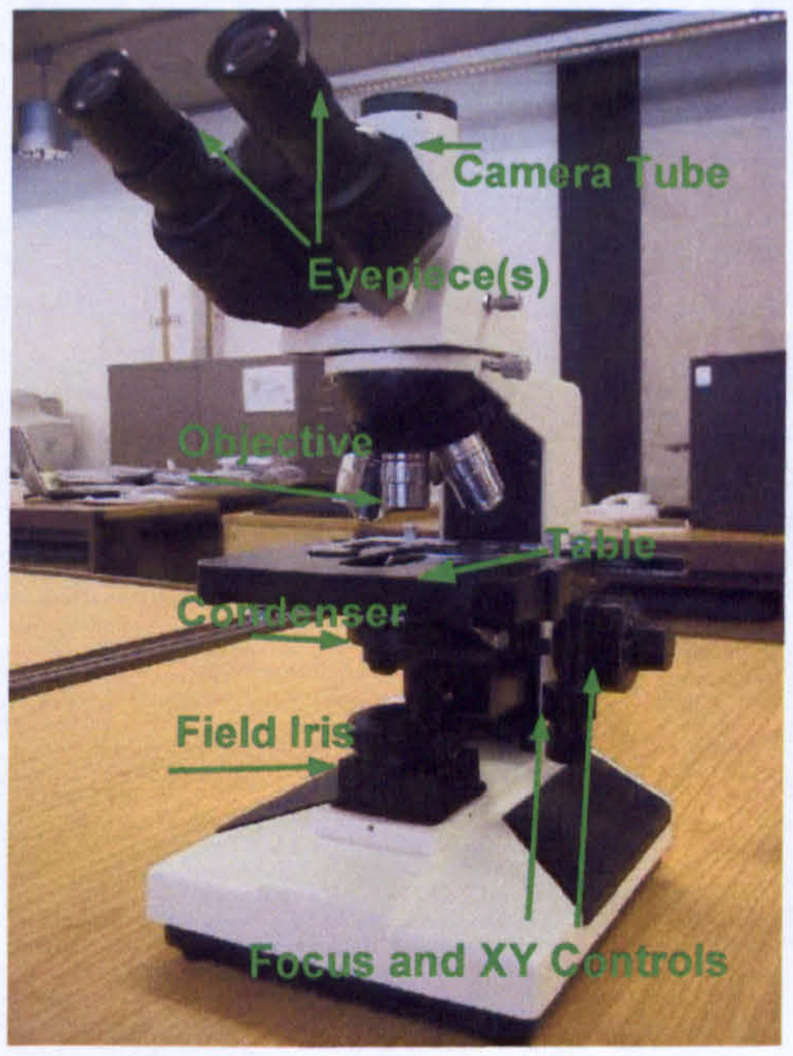
Earlier work by Rao [35] produced software that could read the malaria slide images from disk and process the images. During the earlier stages of this study, we implemented a TWAIN interface for this software to eliminate the necessity of an external process for acquiring images. Currently, this interface can work with both the mini camera and the Canon A60; however, it does not support the full functionality (e.g. remote capture, focus, and colour controls) of the latter.

3.1.1 Quality of the Images

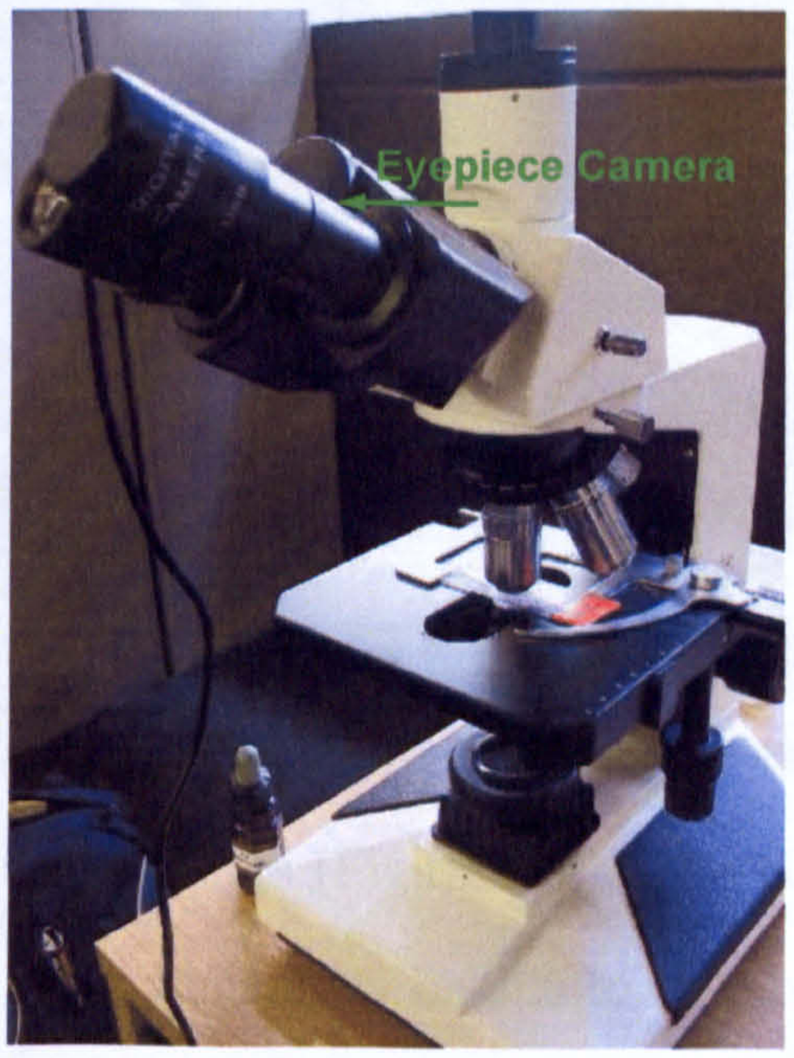
There are several conditions of the acquisition setup that affect the quality of the images. Some of these conditions originate from the quality of the individual components; hence they are difficult to control. However, some of them were related to the calibration of the components.

- The noise and pixel resolutions of the cameras:

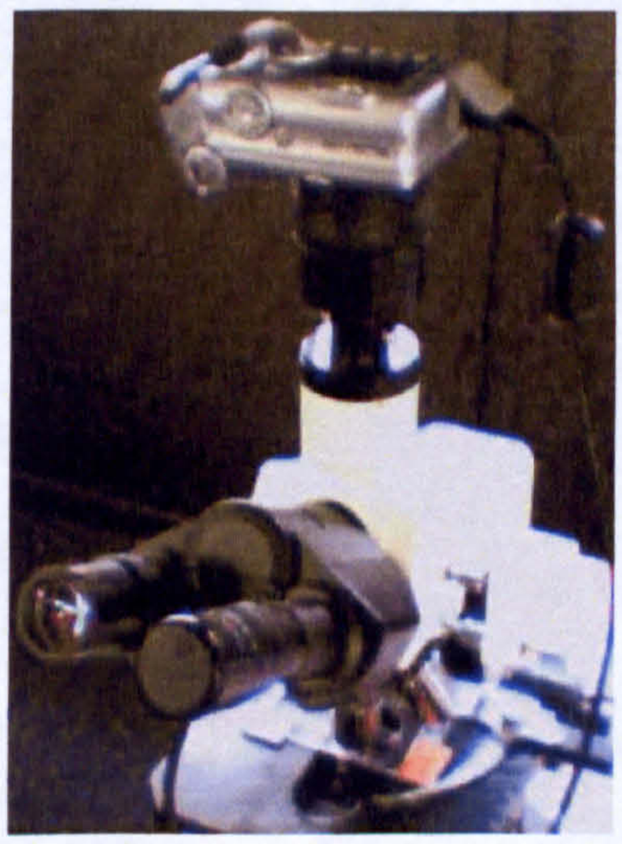
  Camera noise can be related to the various built-in electronic or optic components, but it is mainly due to the quality of imaging sensors. The mini eyepiece (1/3 inch CMOS with 583 x 438 pixels) camera has a significant temporal noise due to its low quality CMOS sensor. In the case of still objects, an intuitive approach which can reduce temporal noise can be an averaging of the successive images. Also, more sophisticated algorithms can be developed based on statistical variation of



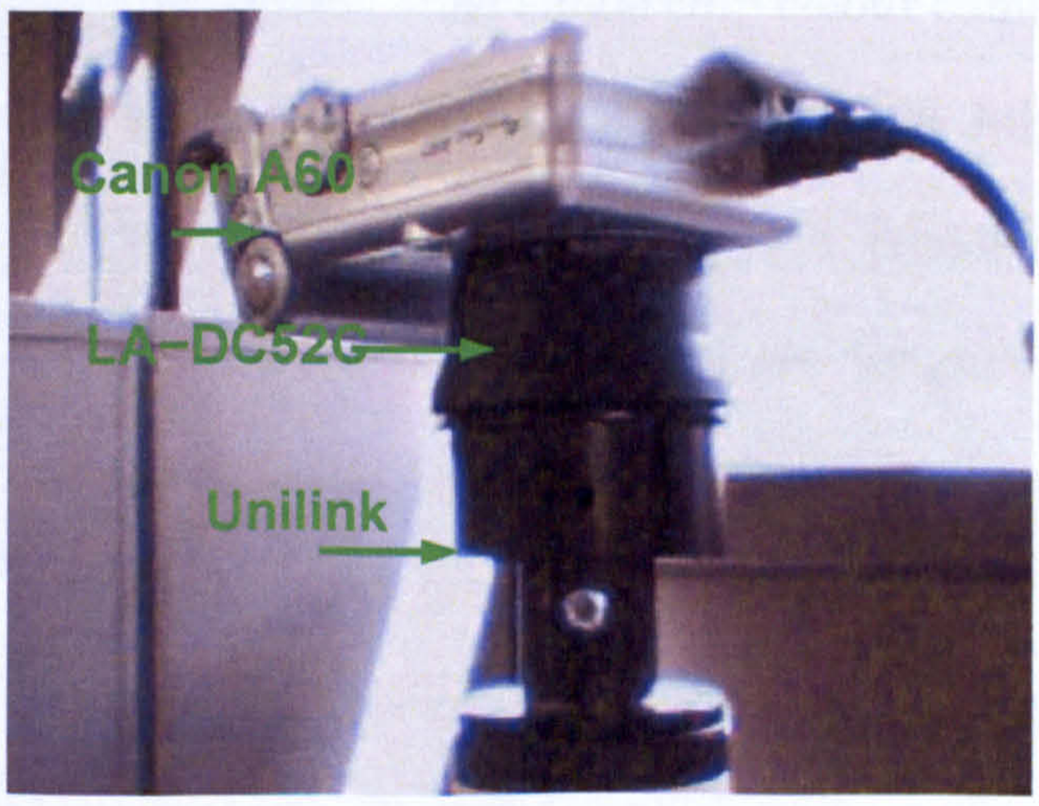
(a)



(b)



(c)

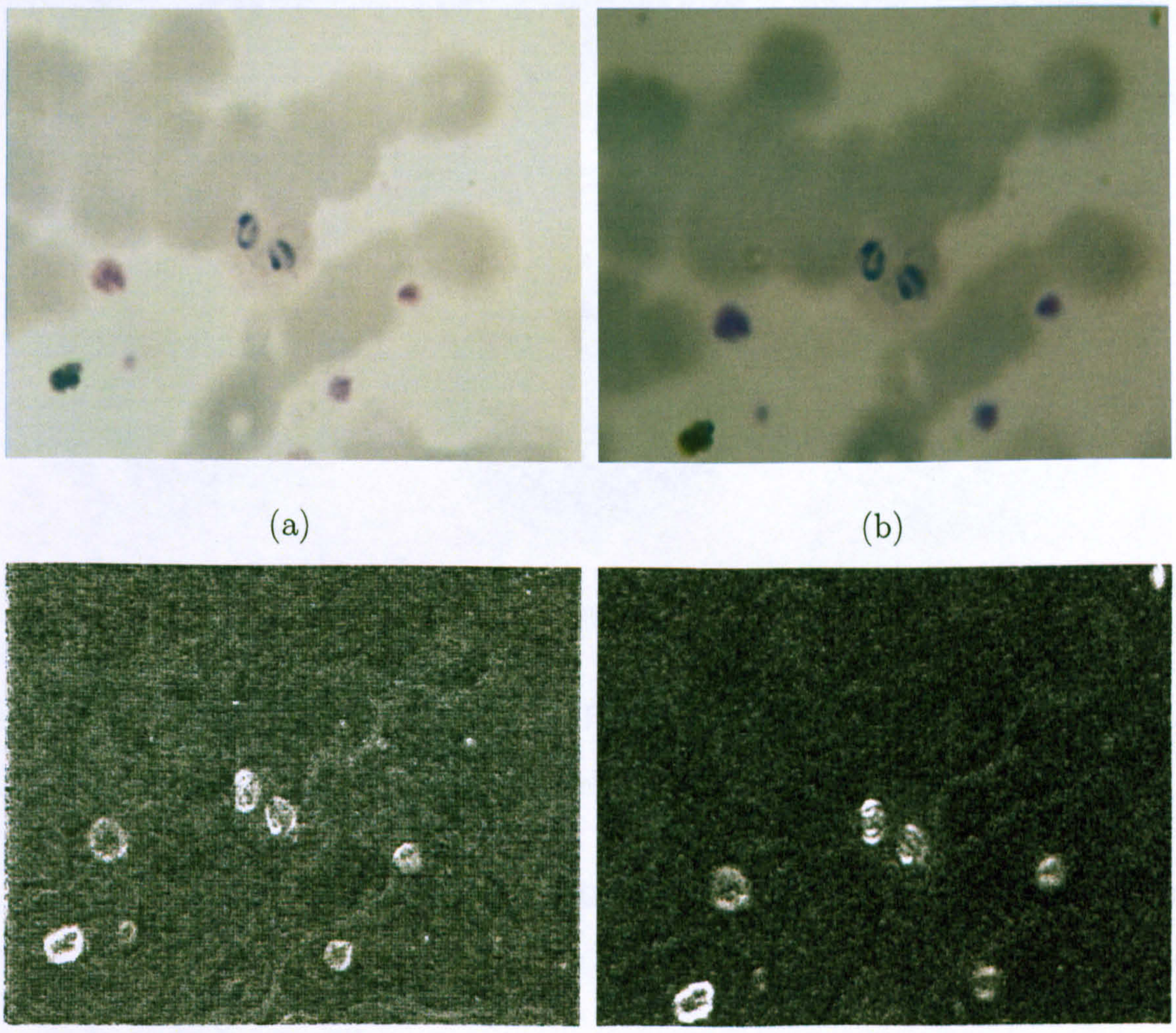


(d)

Figure 3.2: UoW Image acquisition setup: (a) Brunel microscope, (b) Mini eyepiece camera, (c) Canon A60 attached to the trinocular microscope tube, (d) Canon A60, lens adapter and Unilink coupler.

the noise. However, these operations would require additional time. Figure 3.3 demonstrates mean and standard deviation images of ten subsequent images of the same thin film field taken using the eyepiece camera and Canon A60. It can be seen that both cameras have high standard deviation coinciding with the edges of the structures. This may be related to the inevitable slight movements between the fields. In addition, the JPEG compression that is default in Canon A60 may boost this errors. However, the eyepiece camera has a patterned noise structure (see grid lines in the standard deviation image Figure 3.3(c)), and some flawed pixels (bright dots in the middle). Since we had an alternative in the Canon A60, we did not put our efforts to enhance the quality of the eyepiece camera. However, as can be observed on the mean images (Figure 3.3(a) and 3.3(b)) can produce brighter images than Canon A60, which may be due to its direct fitting capability to the eyepiece of the microscope.

Additionally, there is the issue of resolution. Most single chip colour cameras provide the colour image with patented (U.S. patent No. 3,971,065 (20 July 1976)) Bayer colour filter arrays which is named after its inventor Dr. Bryce E. Bayer of Eastman Kodak. In this scheme, individual pixel sensors are designed to filter (and respond to) different colours (i.e. Red, Green, Blue) of the light. Usually the number of colour pixel sensors is not equally distributed among different colours: green is the most populated. Then the colour image is produced by de-mosaicing algorithms which interpolate the actual 3 colour values of the pixels according to the colour values of the neighbour pixels. Due to the use of these algorithms the information on the separate colour channels is usually correlated in single-chip cameras. Moreover, because of the dominant number of green colour pixel sensors, the quality of the green channel is better than the red and blue channels. Hence, the actual resolution of a single chip camera is different among different channels and not more than 1/3 of the overall number of pixels. The best al-



(a) (b)

(c) (d)

Figure 3.3: Temporal noise in the eyepiece and Canon A60 cameras: (a,b) Mean images obtained from ten images taken from the same field using the eyepiece and Canon A60, respectively (c,d) standard deviation images of the same group of images, respectively for the eyepiece and Canon A60. The brighter pixel values indicate higher standard deviation.

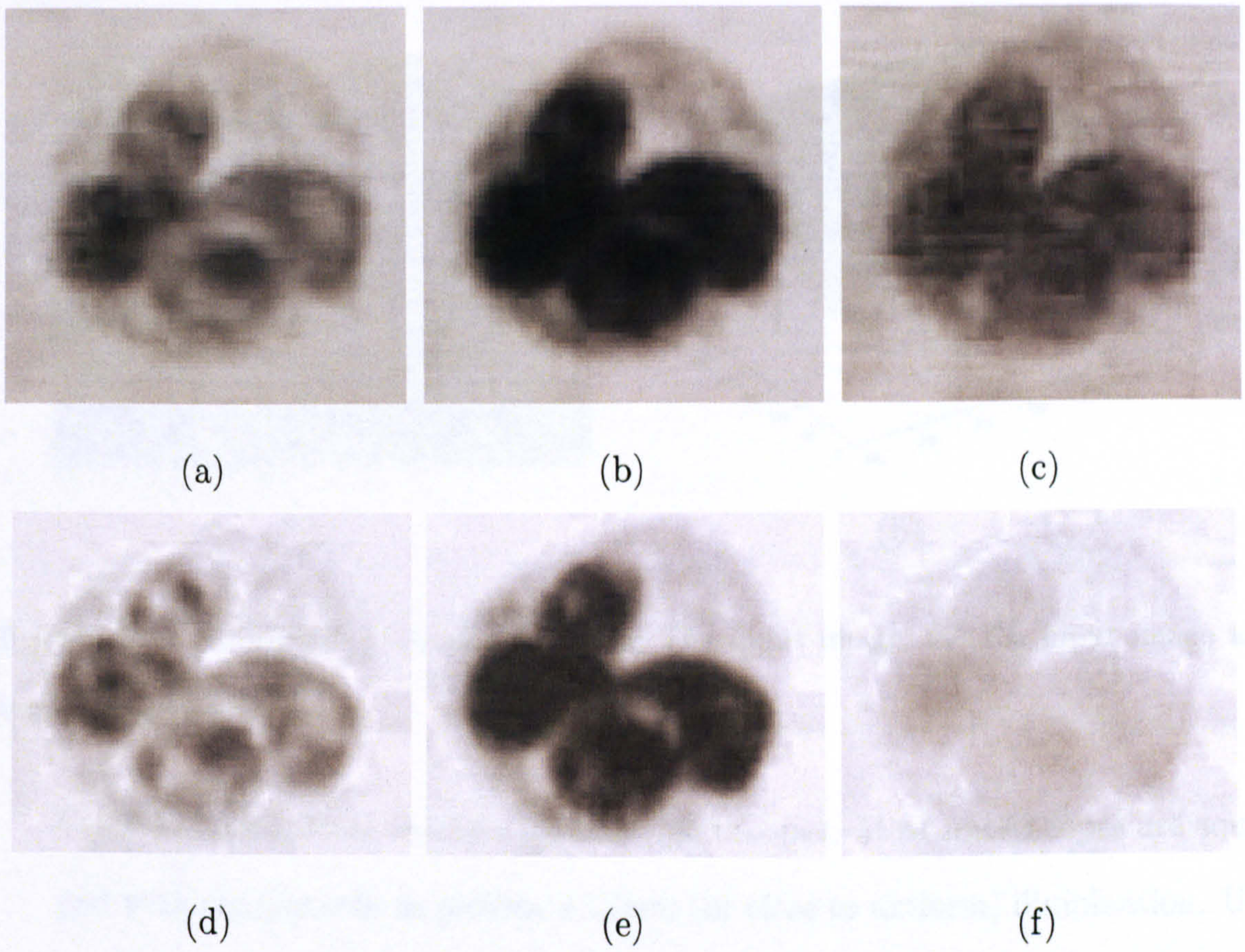
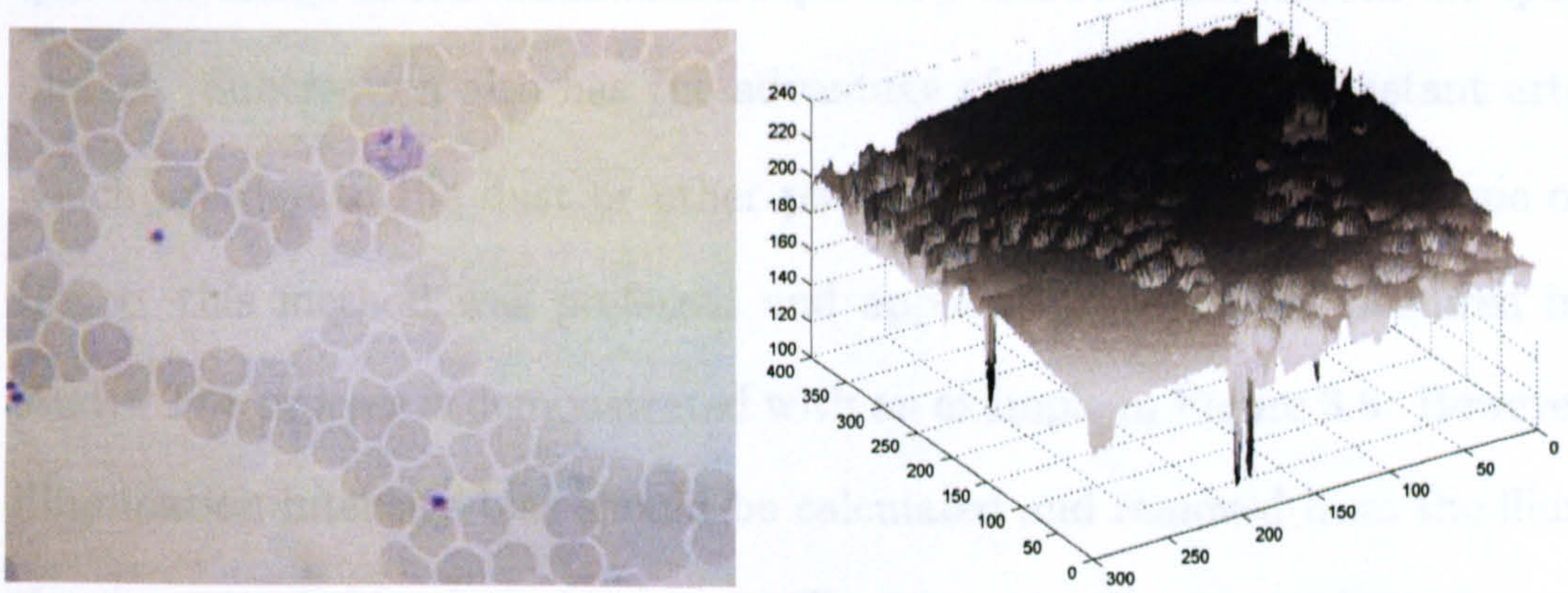


Figure 3.4: Comparison of the cameras: Red, Green, Blue channels respectively from the (a), (b), and (c) eyepiece camera (d), (e), and (f) Canon A60.

ternative is a three chip camera that produces the three separate colours with different associated pixel sensors; it is more expensive however usually preferred in the medical imaging.

Both of the eyepiece camera and Canon A60 have single chip image sensors. Thus, the individual colour channels are highly correlated and actual pixel resolution is less than the overall resolution. We did not measure and compare the noise levels between the cameras but Figure 3.4 visually demonstrates and compares the separate channels (i.e. Red, Green, Blue) of the same thin film field for the two cameras. Clearly the visual comparison is in favour of the Canon A60, i.e. there are less noise and less correlation between colour channels. Therefore the Canon A60 was used for the subsequent experiments.

- Non-uniform Illumination



(a) (b)

Figure 3.5: Non-uniform Illumination: (a) The input image, (b) the input image as topographical surface.

Uniform illumination is important for microscopes. Most microscopes are equipped with components to provide uniform (or close to uniform) illumination. Uniform illumination needs a calibration of the microscope. A common illumination calibration standard is Kohler Illumination named after its inventor August Kohler [113]. In this method transmitted illumination from the light source is aligned and focused for examination of a parallel and uniformly illuminated specimen. Thus, it requires an adjustable condenser height and a variable field diagram (i.e. on top of field iris) on the microscope (as shown in Figure 3.2). This is often neglected by microscopists since the human vision system is adaptive to local illumination changes. In the early stages of this study, though the microscope was equipped to adjust Kohler illumination, because the author was unaware of this method, the results did not benefit from it.

However, for general microscopy imaging and particularly for the CDM application non-uniform illumination can produce serious problems in various stages of the process. Figure 3.5 shows an example colour image and its surface plot to demonstrate the uneven illumination (see the different elevation on the corners). One simple intuitive way of dealing with the non-uniform illumination is to ac-

quire the image of the illumination separately and subtract it from the specimen images. Subtraction also has the advantage of removing the constant artefacts which are due to the dust or other particles located on the microscope optics. Hence, this method was preferred and applied to the images acquired in this study. The process is demonstrated with an example in Figure 3.6. However, the illumination intensity bias should be calculated and removed from the illumination image prior to the subtraction. The bias value for each colour channel can be simply calculated by the first non-zero entry of the corresponding histogram as shown in Figure 3.6(c).

An alternative method is to filter the images to remove the variation in the illumination. If there is a smooth varying illumination in the image, a high pass filtering operation may reduce the potential effects. In mathematical morphology (MM) there is a simple solution to achieve such a filtering: "top-hats" [29]. Depending on the input image condition, black top-hats (for dark objects-bright background) or white top hats (for bright objects-dark background) or both of them can be used (for mixed images, see [29]). For the blood slide images that are obtained with conventional light microscopes, the low pass content can be calculated with closing (on the grey scale) by a sufficiently large (bigger than objects) structuring element. Sufficiently large structuring element size can be determined using granulometry or area granulometry [35]. Then the input grey scale image can be subtracted from the closed image to obtain a high-pass filtered image. In an image with opposite conditions (bright objects-dark background) opening of the image must be subtracted from the input image. For colour images, closing and subtraction operation should be performed in different colour channel images separately. In addition, as in the case of the illumination image subtraction method explained in the previous paragraph, the global bias in each channel can be calculated from the histogram to avoid an undesired darkening

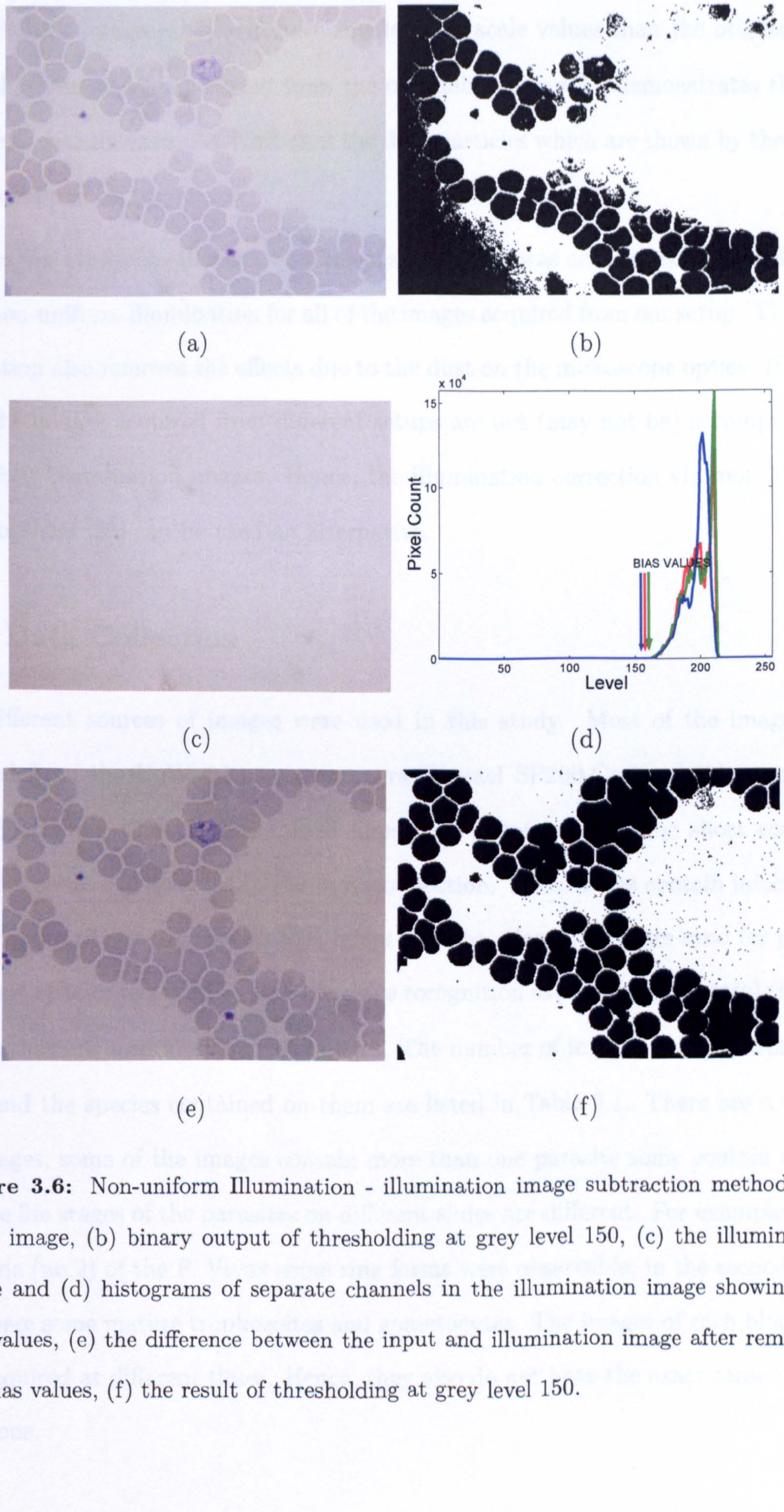


Figure 3.6: Non-uniform Illumination - illumination image subtraction method: (a) Input image, (b) binary output of thresholding at grey level 150, (c) the illumination image and (d) histograms of separate channels in the illumination image showing the bias values, (e) the difference between the input and illumination image after removing the bias values, (f) the result of thresholding at grey level 150.

effect on the images. Note that after removing the bias values from the closed channel images, they will have smaller grey scale values than the originals; and they should be subtracted from the originals. Figure 3.7 demonstrates this process with an example. Note that the dust particles which are shown by the arrows remained on the image.

In this study, the illumination image subtraction was used to reduce the effects of non-uniform illumination for all of the images acquired from our setup. This operation also removes the effects due to the dust on the microscope optics. However, the images acquired from different setups are not (may not be) accompanied by their illumination images. Hence, the illumination correction via morphological top-hats [29] can be used an alternative.

3.2 Data Collection

Two different sources of images were used in this study. Most of the images were acquired from the UoW microscope-camera (Brunel SP200-Canon A60) setup. The slides containing thin or thick blood films were obtained from the short course on Blood Parasites as explained in the previous section. These slides contain infections of all four types of malaria species that infect humans. Hence, they are used for parasite detection, species recognition, and life stage recognition experiments. In total there are 9 slides that are used for acquiring images. The number of images acquired from these slides and the species contained on them are listed in Table 3.1. There are a total of 630 images; some of the images contain more than one parasite some contain none at all. The life stages of the parasites on different slides are different. For example, in the first slide (no.2) of the P. Vivax some ring forms were observable; in the second (no.4) there were some mature trophozoites and gametocytes. The images of each blood slide were acquired at different times. Hence, they also do not have the exact same imaging conditions.

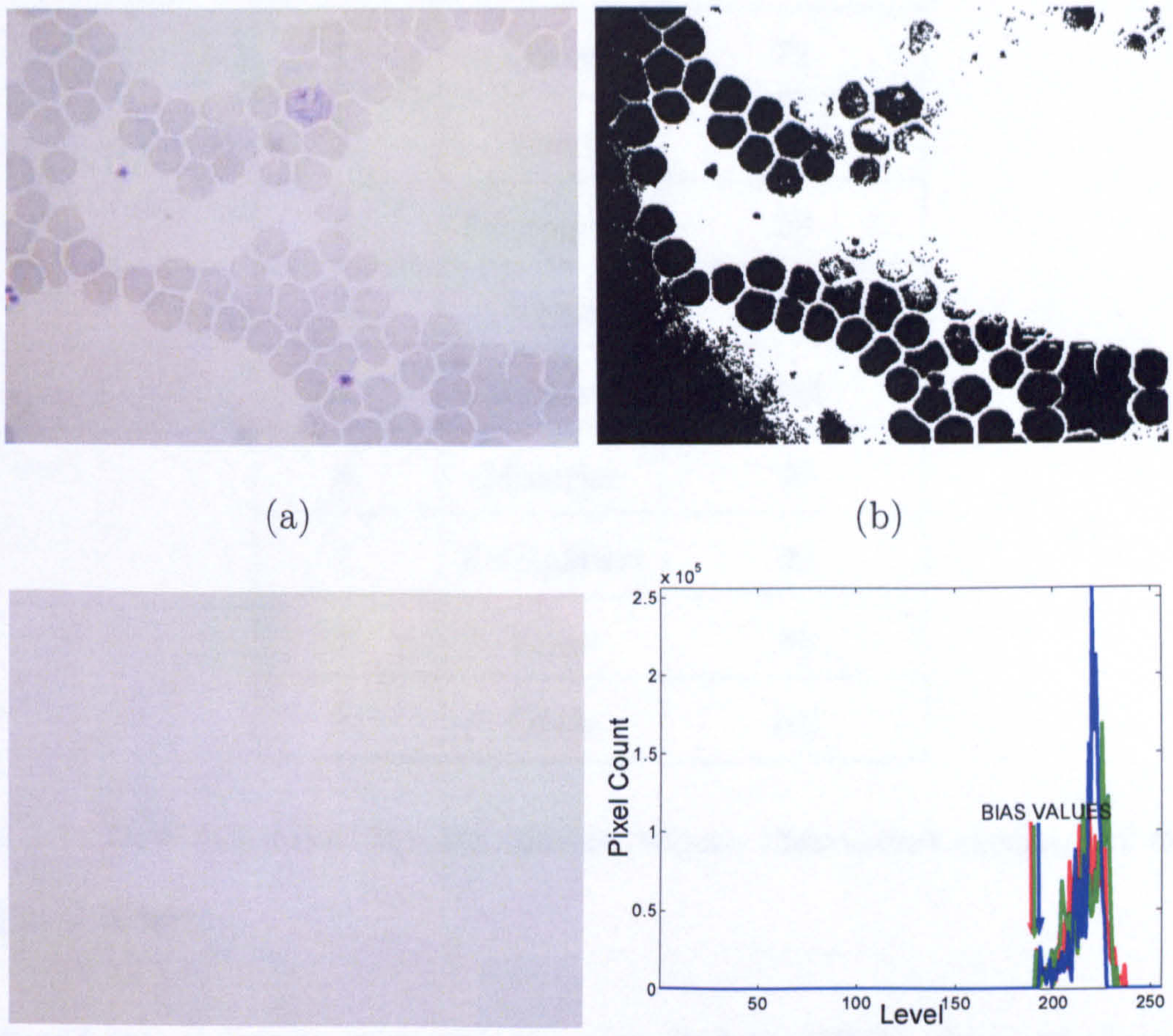


Figure 3.7: Non-uniform Illumination - Filtering with morphological top-hat operation: (a) Original input image, (b) binary output of thresholding at grey level 150, (c) Closing with a large square structuring element, (d) its histograms indicating the global bias values, (e) (c) subtracted from (a) after removing the bias values from separate channel images, and (f) the filtered image thresholded at the same grey level.

Smear #	Species	Num. Images
1.	Ovalc	72
2.	Vivax	92
3.	Falciparum	28
4.	Vivax	70
5.	Falciparum	152
6.	Malariae	40
7.	Falciparum	20
8.	Vivax	56
9.	Ovale	101

Table 3.1: UoW thin blood film set: Smears (slides), *Plasmodium* species, and number of acquired images.

A second set of images were the ones supplied by NIMR. Some of these images contain parasite species that infect only rodents. Parasite species that infect the rodents are different from the parasites that infect humans but they are useful to scientists studying the disease. In this study, these were only used in segmentation tasks. Some others contain culture (i.e. *in vitro* not peripheral blood) examples of *P. Falciparum*. The parasites in the blood culture can be developed under control. Hence, these images also serve research purposes. Parasites in the culture images are similar to the ones in the peripheral blood film images; however Red Blood Cells (RBCs) are slightly distorted. In this study, these images were mainly used in the development and testing of the blood cell image segmentation or colour normalisation algorithms.

As mentioned before, by data we refer to images accompanied by supervised information. The information for an object detection task can simply be the location of the objects (parasites in this case) on the image. In computer vision jargon this information is the *ground-truth* of images; as it is referred to henceforth in this thesis. In

the following section the preparations of different ground-truth data (for the different experiments) are explained.

3.2.1 Segmentation Ground-Truth

Different type of ground-truth information can be created to perform quantitative experiments for the segmentation performance. Since the aim of the segmentation task is to locate individual cell regions, the ideal information can be the manual marked cell borders. However, this requires a significant amount of time and it is difficult to estimate the actual border when two or more cells are touching each other. Simpler information is the cell count. However, the segmentation algorithms extract and count the homogeneous regions. Hence, a performance evaluation can be error prone when the total number of the regions is compared with the total number of cells. An alternative piece of information which is more complex than cell count but requires less labour than marking the cell borders can be the cell centre locations: manually marked locations can be checked if they reside in the automatically found regions. We have used the cell centre location information to assess the performance of the segmentation method. In order to perform manual cell centre marking in the images we have developed a simple program which displays the image and stores mouse clicked coordinates in a text file (Figure 3.8). This program was also intended to assist the manual marking of the parasites and their life stages; however it was not used for this purpose since it was replaced by a more comprehensive program which is explained in Section 3.2.3.

3.2.2 Stained Pixel Detection Ground-Truth

As mentioned before, the Giemsa-stain affects parasites, White Blood Cells (WBCs) and some artefacts. The RBCs are also affected and change their colour; however the stained pixels which belong to the former group are much more obvious (i.e. chromatically saturated) pixels in the image. Hence, the stained pixel detection means

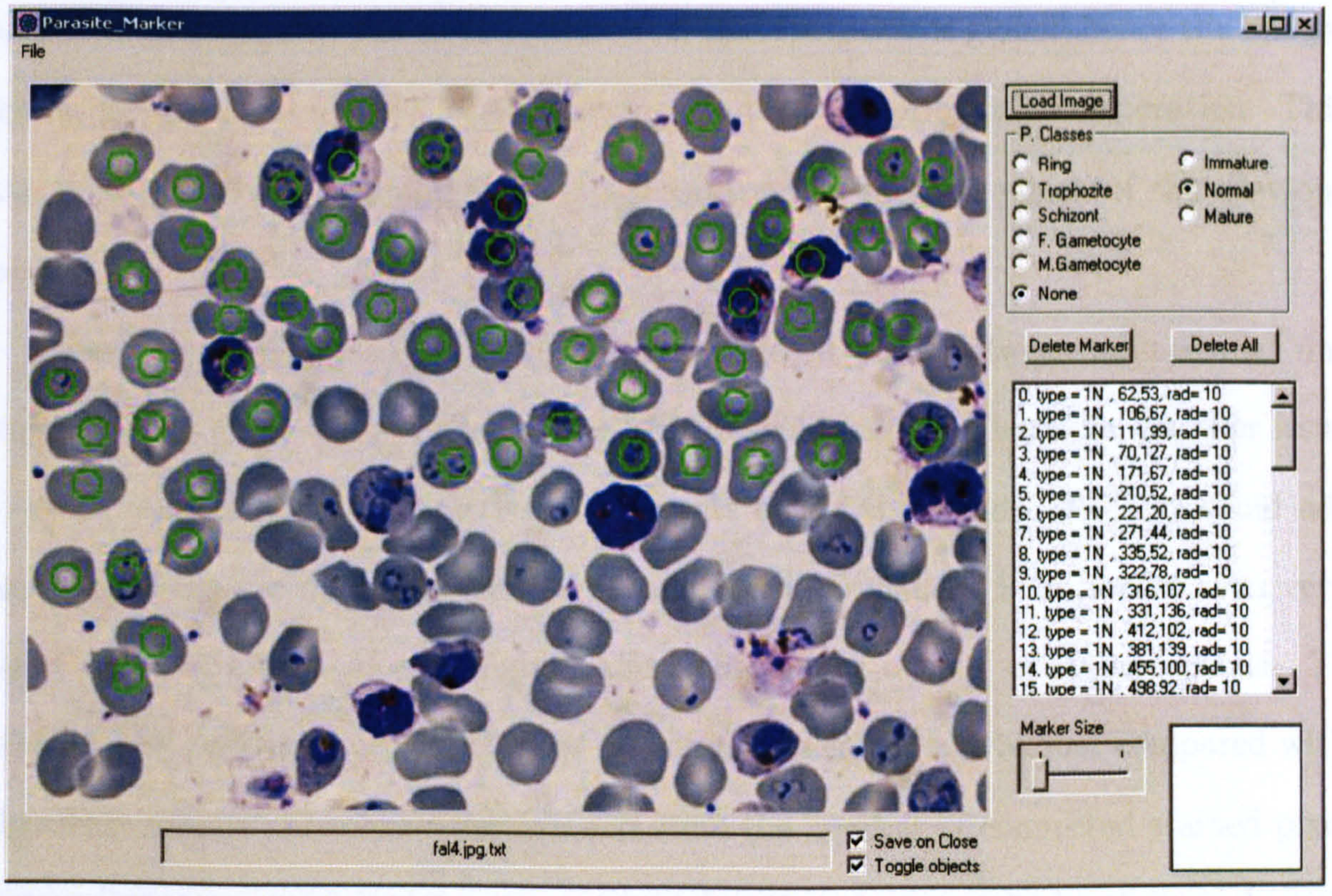


Figure 3.8: The user interface of the segmentation ground-truth preparation program. See the text for details.

detection of these highlighted pixels (pixel groups) in the image. The necessary ground-truth information for stained pixel detection experiments is similar to the segmentation ground-truth. However, in this case we would like to locate only the stained pixels. Marking individual pixels one by one in an average size image (e.g. 640x480) is highly impractical. The alternative information can be the binary masks obtained by manual thresholding operations. In order to provide manually chosen thresholds in every image, we have developed a simple Matlab program with a basic user interface (Figure 3.9(a)). With this program the user can open an image, observe the stained pixels both in the grey level and saturation image (after converting the image to Hue-Saturation-Value colour space) and manipulate two threshold sliders to see the effect of thresholding on the image (Figure 3.9(b)). Two different threshold value choices were provided to perform morphological double thresholding on the grey level image [29]. As explained later in Chapter 4, in this method, the first value performs a rough threshold while the

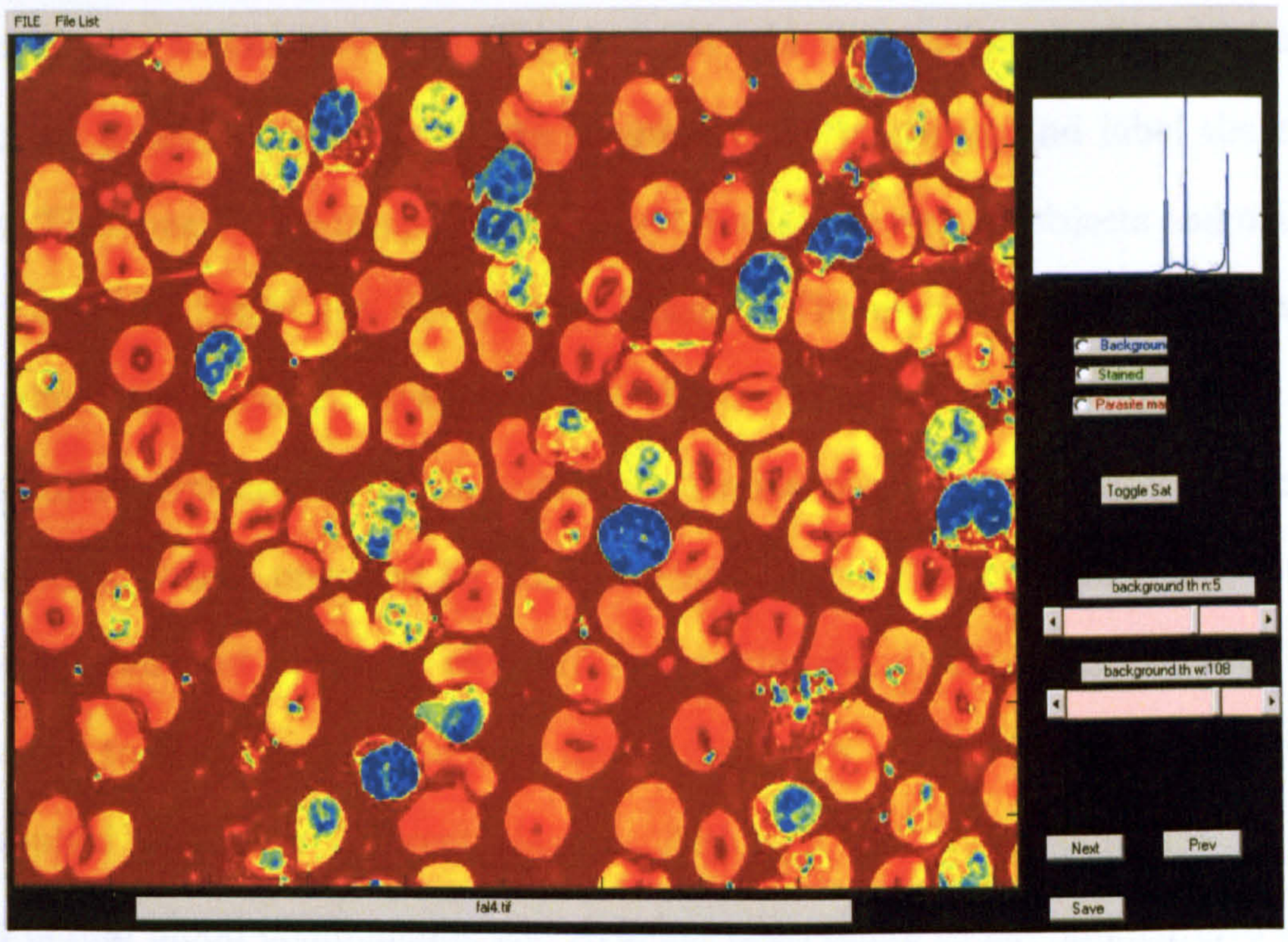
second is used to refine the former. However, this operation is global; hence the images should be corrected against locally varying illumination prior to this operation. This was performed by the illumination image subtraction method which of the previous section.

Since the threshold values operate on a trade-off basis between stained and unstained pixels there is no *perfect choice* for threshold values. Therefore, another issue to point out here is the subjective assessments of the thresholds. The threshold adjustment slides are operating between 0 and 255 grey values. Selection of even very close values of threshold can slightly affect the binary output of stained pixels. To evaluate the performance, the number of stained pixels can not be just compared with algorithm results. Hence, we provide and store the number of connected stained pixel groups (components) together with an image locating the manually located pixels with threshold values.

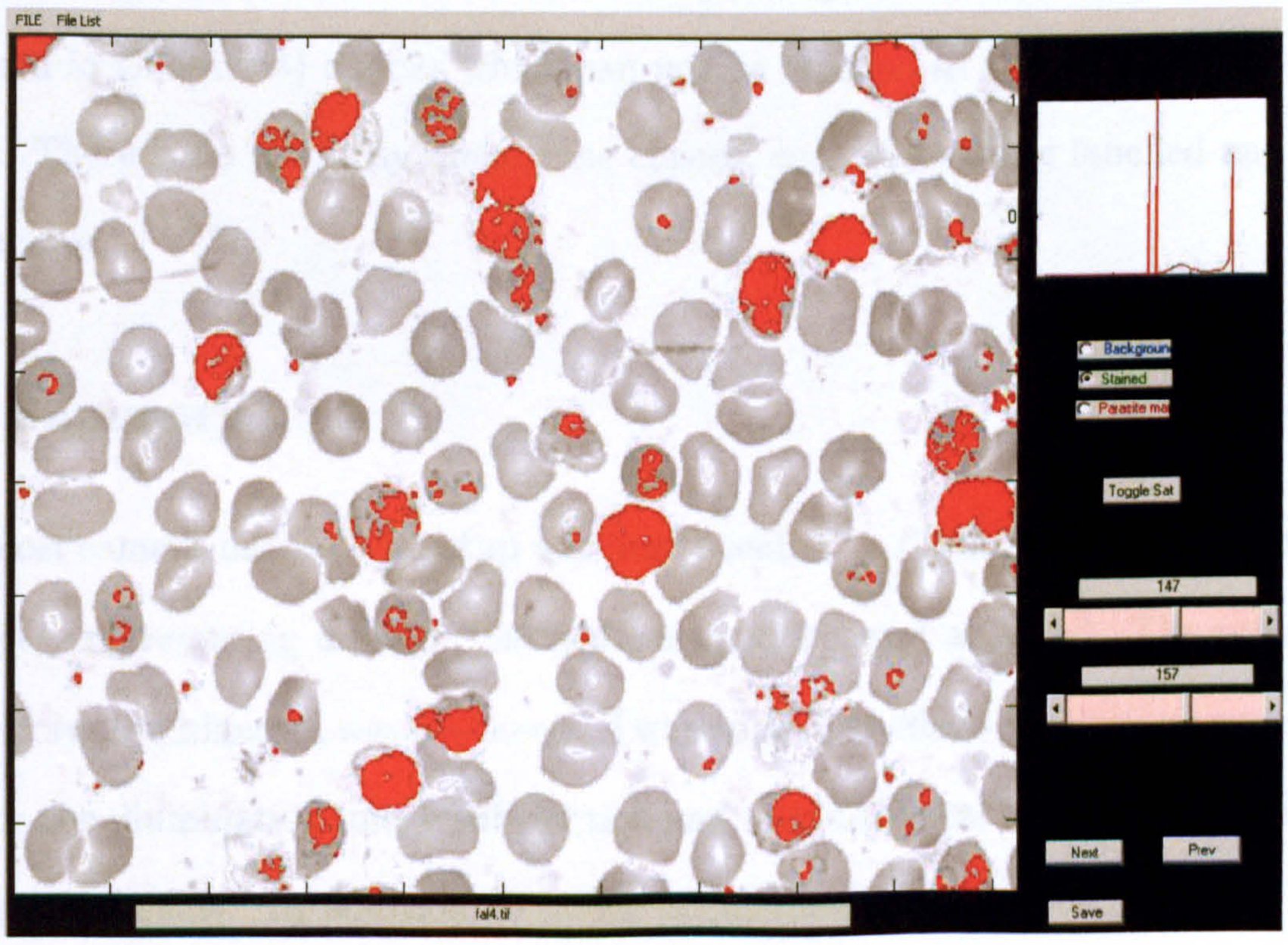
In this thesis, we propose a stained pixel detection method based on the conditional density distributions and Bayesian decision making (see Chapter 6). Therefore the method requires training samples to estimate the density distributions. By storing manually thresholded images which mark the stained pixels we provide data to be used in both training and performance evaluations.

3.2.3 Parasite Detection Ground-Truth

The automatic parasite detection algorithm takes a stained object in the image as an input and decides whether it is a parasite or not. In order to develop and evaluate the parasite detection algorithm, we have to identify and locate the parasites. However, we have not directly taken the images and labelled the parasites and other objects. To provide coherent information with the preceding tasks such as stained pixel detection, we have developed a simple program to supervise their outputs (candidate stained objects). Thus, we have aimed not only to show the parasites location, but to observe



(a)



(b)

Figure 3.9: The stained pixel ground-truth preparation program: (a) user interface, (b) output stained pixel map after manual selection of the thresholds.

and supervise potential inputs to the parasite detector. For this task, we have developed another simple Matlab program (Figure 3.10). In this program, after an image is loaded, the program performs all the required tasks to obtain and label the stained objects. Then one by one the program shows (marks) the stained objects and magnifies them (Figure 3.10 right top and bottom images). The user is asked to provide (by a mouse click) the correct category of the object (Figure 3.10 left top). Since, this task requires identification of the parasites among other stained objects, there are corresponding classes provided for non-parasite objects. The program is also intended to serve species and life-stage recognition experiments. Hence, the parasite classes are divided to four species and four life-stages under them. The non-parasites are also supplied with four different classes. Although the *WBC* and *Platelet* classes refer to actual blood components, the *Artefact* class refers to any unidentified stained object; the last non-parasite class *Partial* designates partially found (by the algorithm explained in Chapter 6) regions which can not be labelled as a whole object for other classes. There is no hierarchy among the classes, each class being labelled and stored individually.

3.3 Summary

A low-cost camera-microscope setup was implemented (~£1000). A significant number of images (representing all four *Plasmodium* species) were acquired. The problem of non-uniform illumination was discussed. Two simple and effective solutions were implemented: the illumination image subtraction and morphological top-hats based filtering as proposed in [29]. In addition to image acquisition, the ground-truth data collection has been explained. Three programs were implemented to aid manual markings which are used in the segmentation, stained pixel detection, and object classification experiments.

Figure 3.10: Parasite detection ground-truth preparation program: the user interface of the program showing (first column) the class selection options, input image and (second column) binary mask locating the stained object, and zoomed stained object on colour image.

Chapter 4

Segmentation

4.1 Introduction

Probably one of the most common shared tasks in image analysis systems is segmentation. Segmentation aims to partition the image plane into meaningful regions. The definition of the meaningful regions and partitioning method is usually application specific. For example, the methods can be aimed at separating foreground-background, moving-still regions or objects with specific properties. The segmentation strategy can be a hierarchical partitioning that operates deductively to define first a higher level of object plane, then sub-structural components of the objects or the opposite: as it first defines objects of interest with a specific property then it performs higher levels of partitioning.

In Computerised Diagnosis of Malaria (CDM) either inductive or deductive approaches can be followed. In Di Ruberto's research [34] first the stained objects were identified by their colour properties; then only the Red Blood Cell (RBC) regions containing the stained objects were segmented. In Rao's research [35] the deductive strategy was followed: the image was first segmented into foreground and background regions; then foreground regions were segmented to obtain individual RBC regions; then RBCs were further segmented for parasite analysis.

In the early stages of this study, it was planned to extend CDM research based on Rao's segmentation method. The initial assumption was that the segmentation algorithm was providing satisfactory results for further analysis. However, after some results were obtained from several new input images, it was observed that the algorithm was sometimes produced over-segmented regions, particularly for stained objects. Hence, further study of the segmentation process was required and needed novel development.

Avoiding repeat of the literature review in Chapter 2, the relevant literature can be briefly summarised here. It was explained before that the term "blood cell segmentation" has been used to refer to different notions in the literature. In White Blood Cell (WBC) analysis studies such as [84, 88, 114], "blood cell segmentation" refers to the localisation of WBCs and segmentation of the WBC into the structural components such as cytoplasm and nucleus. Here, the "segmentation" refers to the partitioning of a peripheral thin blood film image of mainly RBCs into separate labelled regions each representing an individual blood component for further analysis.

Some methods exist which are directly related to blood cell image segmentation as described in this study [26, 34, 35, 73, 94, 95]. The common problem associated with this task is the under/over-segmentation of the cells. Under-segmentation is usually caused by unresolvable cell boundaries of contacting or overlapped blood cells. Over-segmentation on the other hand has several causes. They can be related to inhomogeneity of the cell region or algorithm specific problems. Several techniques have been proposed for segmentation and preventing under-segmentation: granulometries and regional extrema analysis [34, 73]; distance transform and area tophats [35, 95]; Bayesian colour segmentation and watershed segmentation improved by discrete chamfer distance [94]. However, most of the studies do not provide detailed evaluation of the segmentation performance nor do they solve the under/over-segmentation problems completely. In blob segmentation tasks, a common technique to reduce under-segmented regions is to utilize the distance transform on the binary image [29]. In [95], this method was utilised in an iterative method to reduce the under-segmented regions.

In this study, a new thin blood film image segmentation method is presented. Figure 4.1 shows the flowchart of the algorithm. The method defines a new MM tool, *minimum area watershed transformation* (MAWT) and applies it to perform an initial segmentation using area information. This modified watershed transformation introduces a solution for the over-segmentation problem associated with the classical watershed algorithm [47]. It eliminates the necessity of marker calculation prior

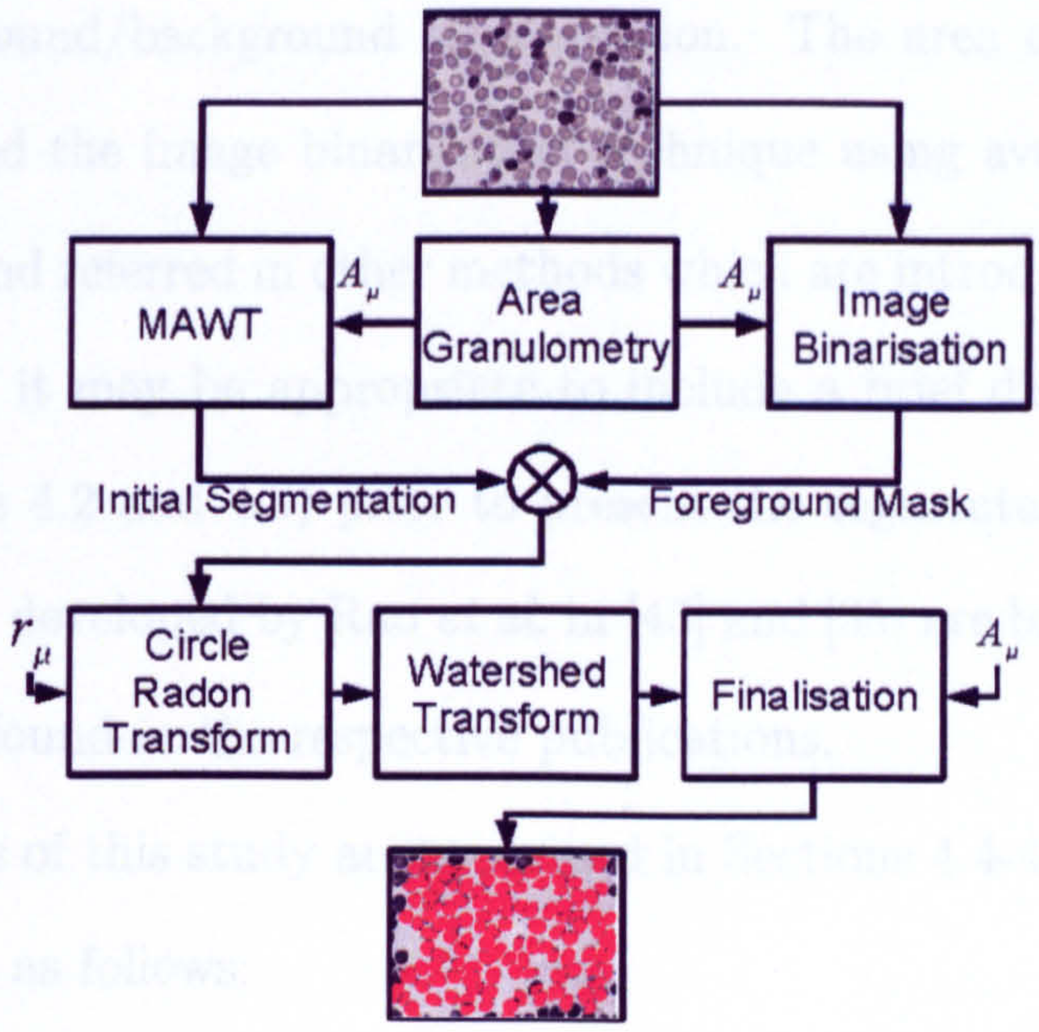


Figure 4.1: The proposed segmentation algorithm.

to watershed segmentation. However, the initial segmentation result contains under-segmented regions. To reduce the number of these regions, cells (actually cell centres) have to be located and a new segmentation must be performed. For this task, a circular object centre locater algorithm has been derived using the Radon transform [96]. The final result is produced by applying the marker controlled watershed transform [29] with markers obtained from the circle Radon transform. The proposed method can be applied to similar blob object segmentation problems by adapting RBC characteristics to the new blob objects.

In order to assess the method's segmentation performance, a benchmark test set with segmentation ground-truth data was prepared as explained in Chapter 3. The method has been tested on the benchmark set and scored a successful correct segmentation rate of 93.7%. In addition, the segmentation performances of the proposed algorithm and the earlier algorithm by Rao *et al.* [35, 95] were compared.

Most of the steps in the proposed method use average blood cell area as an attribute (Figure 4.1). The average cell area is estimated using area granulometry (area pattern spectrum) [45]. The method developed by Rao [35] for the image binarisation is also

used here for foreground/background segmentation. The area estimation with area pattern spectrum and the image binarisation technique using average cell area information are utilised and referred in other methods which are introduced throughout this thesis study. Hence, it may be appropriate to include a brief discussion about these methods (in Sections 4.2 and 4.3) prior to present the segmentation method. These methods, which were developed by Rao *et al.* in [45] and [35] are briefly explained here, more details can be found in the respective publications.

The contributions of this study are explained in Sections 4.4-4.7. The main contributions can be listed as follows:

- A new MM tool -minimum area watershed segmentation- is developed.
- A new marker extraction method based on the Radon transform is developed.
- These two tools are applied to the thin blood film segmentation problem.
- The performance of the overall segmentation method is evaluated on a manual marked test set.
- This study was presented in International Symposium On Mathematical Morphology (April, 2005) and published in the proceedings [40].

4.2 Cell Area and Radius Estimation

A literature review about the size estimation was presented in Section 2.2. In [75], Breen and Jones extended the definition of granulometry to be calculated with any set of attribute openings or non-increasing opening operations. To cover deformed non-circular shapes of cells, Rao *et al.* [45] proposed the utilisation of area granulometry instead of granulometry. In this thesis, many of the proposed methods utilise area granulometry to obtain the size distribution which is used for average cell area estimation. Moreover, in the classification experiments, area granulometry is employed as a shape descriptor feature (Section 7.2.3).

The area granulometry of a grey level image can be calculated as follows:

$$G_\Lambda(X) = \sum \gamma^a_{\lambda_i}(X) - \sum \gamma^a_{\lambda_{i-1}}(X) \qquad \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\} \tag{4.1}$$

where $\gamma^a_{\lambda_i}(X)$ is area opening of image X with area threshold λ_i. See Section 1.8 for a description of area opening. Since area opening operates on the bright regions in the image, the objects of the interest should have brighter grey values than the background. For the images with the opposite properties (i.e. objects dark, background bright) area granulometry can be calculated using the negative image.

The blood cells appear in the grey level image (I) as darker, curved and convex regions (Figure 4.2(a)). Calculating area granulometry (area pattern spectrum) on the negative of I (Figure 4.2(b)) gives a good estimate of the average cell area (A) and the pseudo-radius ($r = \sqrt{\frac{A}{\pi}}$) [45]. Figure 4.2(c) shows area distribution of the bright objects in the negative of the image in Figure 4.2(a). It can be seen that area granulometry plot has a peak at a particular area index. This area index A_1 (and pseudo-radius r_1) is due to the RBCs since most of the image area is covered by the RBCs of similar size.

The area granulometry is a powerful tool to extract size of the objects, however, it has limitations. As explained in Section 2.2, in some anomaly conditions (e.g. anisocytosis, sickle cell) RBCs have irregular shapes (e.g. areas, radius). In Figure 2.1 an example of this condition was shown. In highly overlapping cell fields or in conditions that cause irregular size/shape of cells the area granulometry function still reflects the area distribution of the image however estimating the average cell area (using the maximum index) can be difficult. Simply, there can be other peak values (due to WBCs, enlarged RBCs due to the parasites, and touching RBCs with very weak boundaries, e.g. see Figure 4.6) and thus the pseudo-radius calculation can produce poor results (Figure 2.2). The image binarisation technique and the MAWT algorithm utilise only the area information; however the circle Radon transformation uses pseudo-radius.

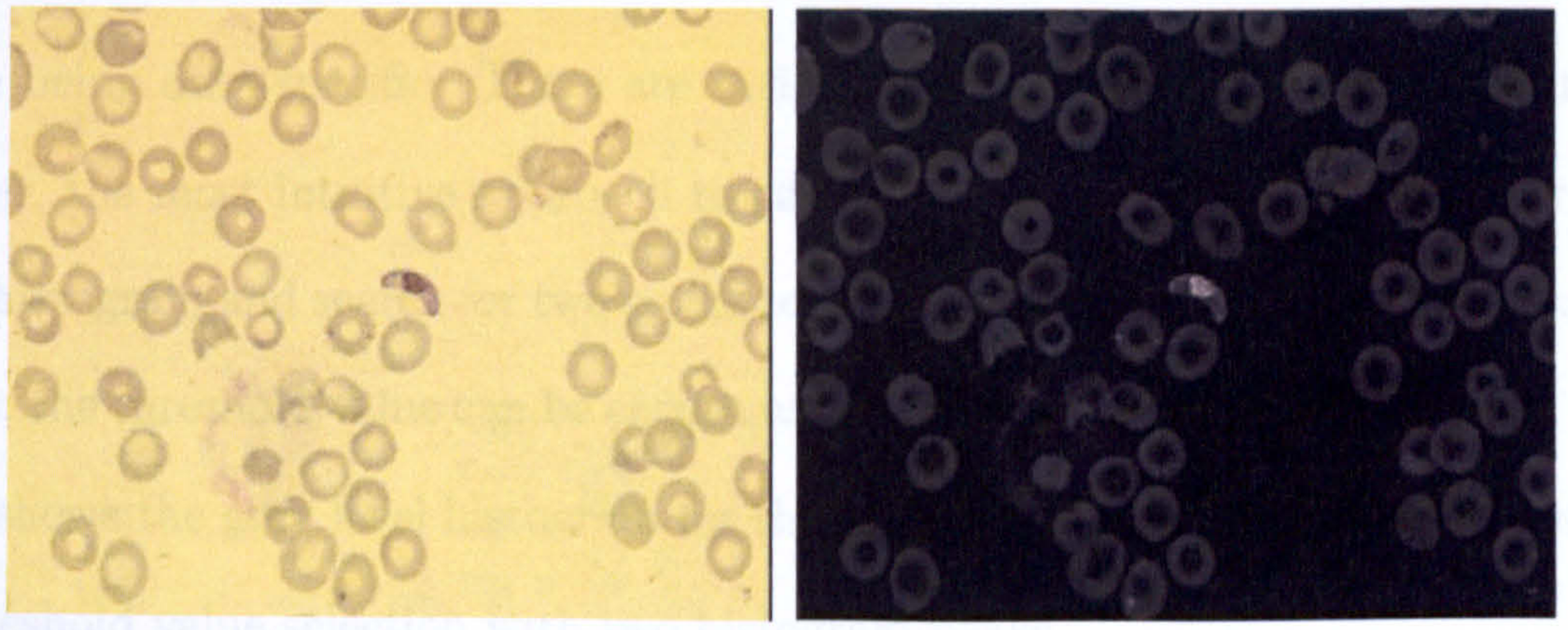


Figure 4.2: The area granulometry calculation for a grey level image: (a) Input RGB image (I), (b) grey level negative, (c) area granulometry and peak area A_1, (d) grey level histogram.

4.3 Image Binarisation

The aim of image binarisation is to partition an image into two separate regions. For the processing of blood cell images, initially, foreground and background regions must be separated. The simplest approach to perform this task is thresholding. However, the threshold value should be automatically calculated to avoid human intervention, despite being data-specific. There are various methods to perform thresholding [85]. However, the most intuitive approach to calculate threshold value(s) is to extract the image histogram and search for two peaks corresponding to foreground and background regions; the threshold value can be chosen as the minimum point between them. Figure 4.2(d) shows the grey level histogram for the image in Figure 4.2(a) and demonstrates the threshold value selection with two histogram peaks.

When the image histogram is not bimodal (as in Figure 4.3(a), 4.3(b), and 4.3(c)) finding two peak values is not easy. However, in thin blood film images foreground and background segmentation is usually possible and efficient with MM operations using the average area of the cells. The thin blood film thresholding method developed by Rao [35] utilised average cell area information to extract a rough estimation of foreground and background regions; then threshold values for foreground and background regions were found from separate histograms calculated from these regions' pixel values. This method is employed to obtain foreground/background segmentation in this thesis. Here, it is briefly explained with a new simpler formulation. The original procedure can be found in [35].

Using cell area information an initial foreground/background separation can be calculated by area opening operation (see Section 1.8):

$$B = (\gamma^a_{(A_1-\epsilon_1)}(I_{ng}) - \gamma^a_{(A_1+\epsilon_2)}(I_{ng})) > 0 \tag{4.2}$$

where two different area openings with $A_1 - \epsilon_1$ and $A_1 + \epsilon_2$ represent less and more than the estimated average cell area value, respectively. $A_1 - \epsilon_1$ value should eliminate very small structures on the image which may not be related with cells, whereas $A_1 + \epsilon_2$

bounds the area opening threshold with a value slightly higher than the estimated cell area value. The area openings are performed on the negative of the grey level image (I_{ng}). Then the non-zero difference provides a rough binary estimate of foreground objects. If B is then used as a mask on the grey level image (I_g), the two separate foreground and background images and their histograms can be obtained. Let us denote the pixel-wise multiplication operator as .* and the histogram operation on a general image as $H_I \hat{=} \|I\|$, then the calculation is as follows:

$$I_{fg} = I_g .* B \qquad I_{bg} = I_g .* B^c$$
$$h_{I_{fg}} \hat{=} \frac{\|I_{fg}\|}{\sum(B)} \qquad h_{I_{bg}} \hat{=} \frac{\|I_{bg}\|}{\sum(B^c)} \tag{4.3}$$

where B^c denotes B complement in the image plane, and $h_{I_{fg}}$ and $h_{I_{bg}}$ are the normalised histograms for the foreground and background regions. Figure 4.3(c) and 4.3(d) show the histogram of the whole image (I, h_I) and the foreground mask (B), respectively. The foreground (I_{fg},$h_{I_{fg}}$) and background (I_{bg},$h_{I_{bg}}$) grey level images are shown in Figure 4.4(a) and 4.4(b), respectively. In Figure 4.4(c), it can be seen that peak indices ((p_{fg}, p_{bg})) for the foreground and background histograms are clearly separated. A "pseudo-histogram" [35] can be obtained by merging the two individual histograms:

$$h_e = argmax(h_{I_{fg}}, h_{I_{bg}}) \tag{4.4}$$

A threshold that is calculated (e.g. $t = (p_{fg} + p_{bg})/2$)) from these two peak indices can be used directly to obtain a binary image (i.e. $T = I_g > t$). However, in MM there is a more efficient thresholding technique: the morphological double threshold [29]. For a grey level image (I_g) this method is described as follows:

$$TT_{[t_1 \le t_2 \le t_3 \le t_4]}(I_g) = R^{\delta}_{T_{[t_1,t_4]}(I_g)} \left[T_{[t_2,t_3]}(I_g) \right] \tag{4.5}$$

where T is the threshold, R^{δ} is the morphological infinite reconstruction operator [29], and $[t_1, t_4]$, $[t_2, t_3]$ are the wide and narrow threshold intervals respectively.

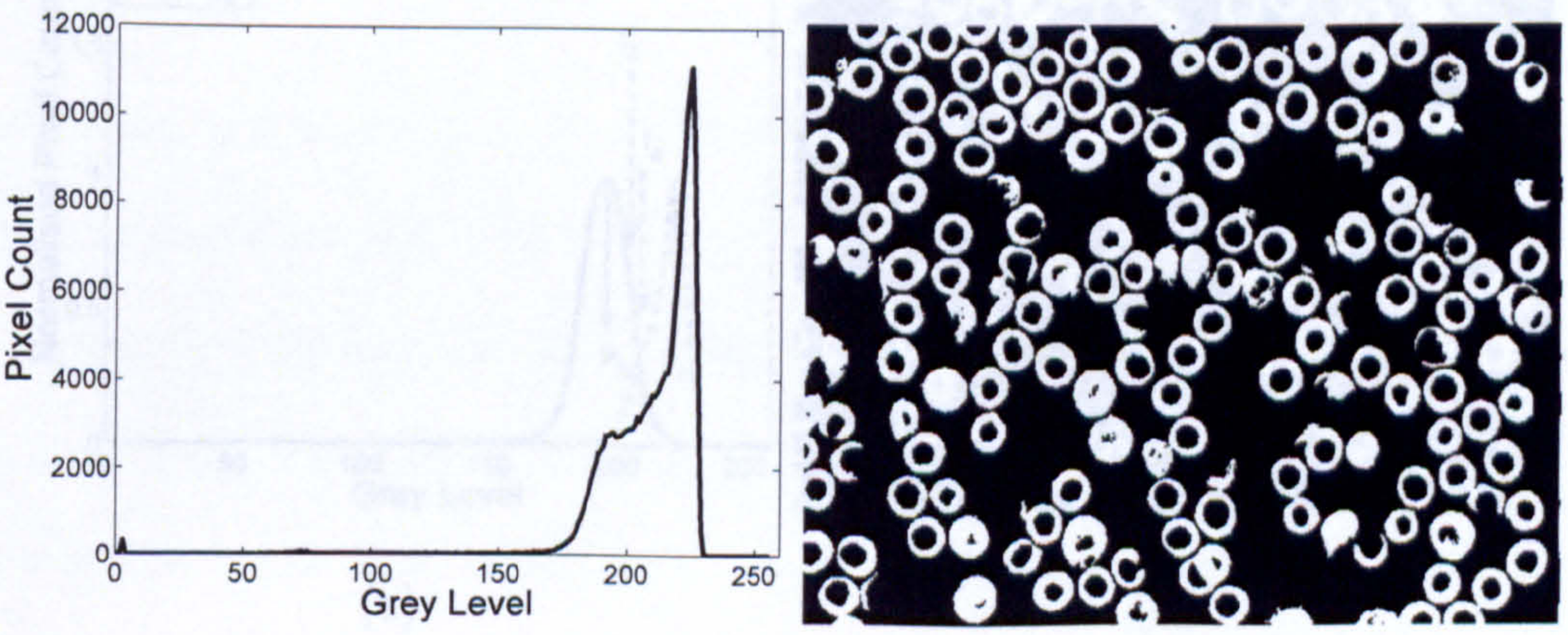


Figure 4.3: Image binarisation - foreground mask: (a) Input RGB image, (b) grey level image I_g, (c) grey level histogram, (d) foreground mask (B)

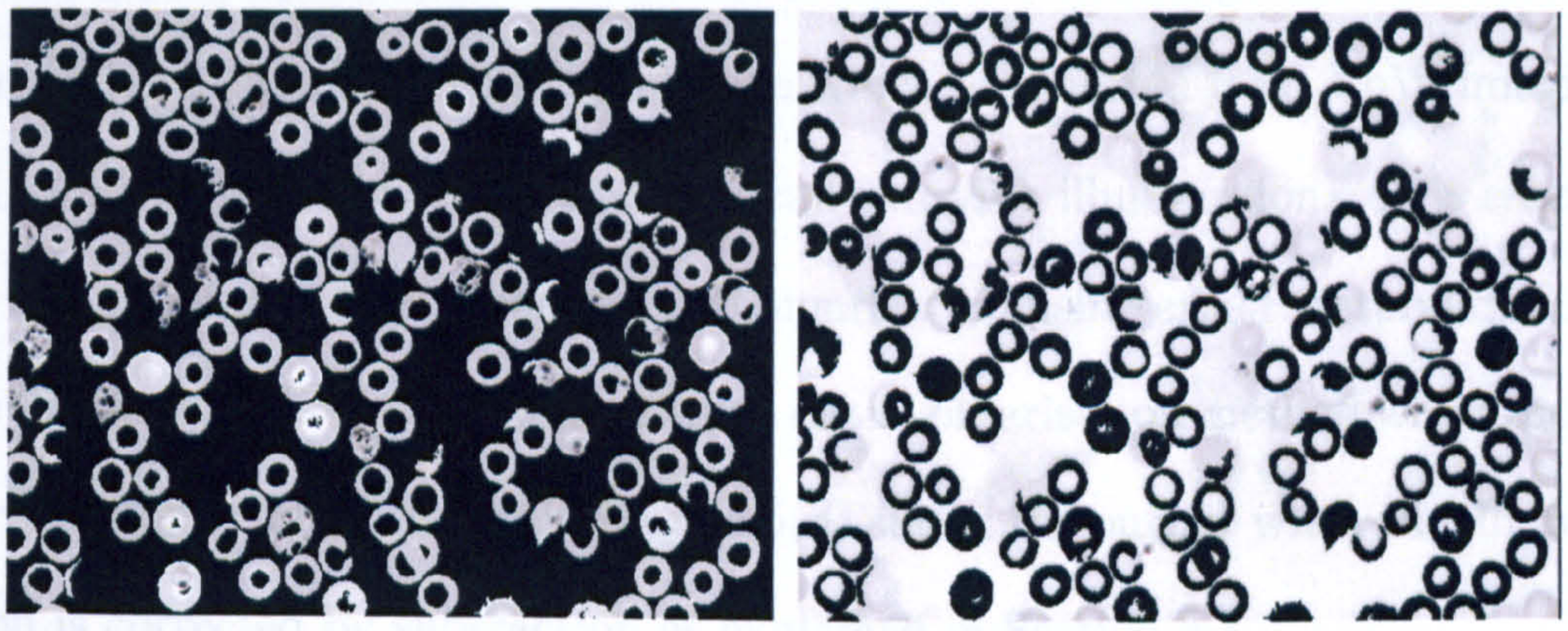


Figure 4.4: Image binarisation - double threshold: (a) Foreground grey level image I_{fg}, (b) background grey level image I_{bg}, (c) foreground and background histograms $h_{I_{fg}}$, $h_{I_{bg}}$ and peak indices $((p_{fg}, p_{bg}))$, (d) Output binary image from double threshold $TT_{[t_1 \leq t_2 \leq t_3 \leq t_4]}(I_g)$

A refined binary representation of the image foreground (output of the binarisation process Figure 4.4(d)) can be obtained when TT is calculated on the I_g with threshold values $t_1, t_2 = 0, t_3 = (3 * p_{fg} + p_{bg})/4, t_4 = (p_{fg} + p_{bg})/2$ (Figure 4.4(c)).

However, contrary to its original aim and presentation in [35], due to the global thresholding (4.5) operation even this method is not able to cope with local variations in illumination. Figure 4.5 shows an image which is acquired from UoW imaging set-up and its histogram (low contrast and locally varying illumination). It is essential to correct the illumination variation prior to binarisation using either method discussed in Section 3.1. Figure 4.5(c) shows the output of the binarisation method when the illumination variation is not corrected; Figure 4.5(d) shows the output when the illumination variation is corrected by subtraction as explained in Section 3.1.

4.4 Minimum Area Watershed Transformation

The watershed transformation is a powerful morphological image segmentation tool initially introduced by [115]. In the watershed transformation the input image is seen as a topographical surface. In the output, watershed lines divide the image plane into regions associated with regional minima. There are several watershed algorithms in the literature [47, 115, 116] which differ in the flooding realisation and computation. However, common to all is that every regional minimum will be associated with a unique region (label) in the output. Applying a watershed transform on the image directly is generally useless unless the objects are flat or at least smooth grey level regions. Hence, a marker controlled transform is usually preferred which basically replaces the regional minima with the externally supplied markers. This technique transforms the watershed segmentation problem to one of marker extraction [115]. Some marker extraction techniques can be found in [29, 117, 118].

The minimum area watershed transformation (MAWT) is a modification to the classical watershed transformation [29, 47], which ensures the areas of the labelled regions

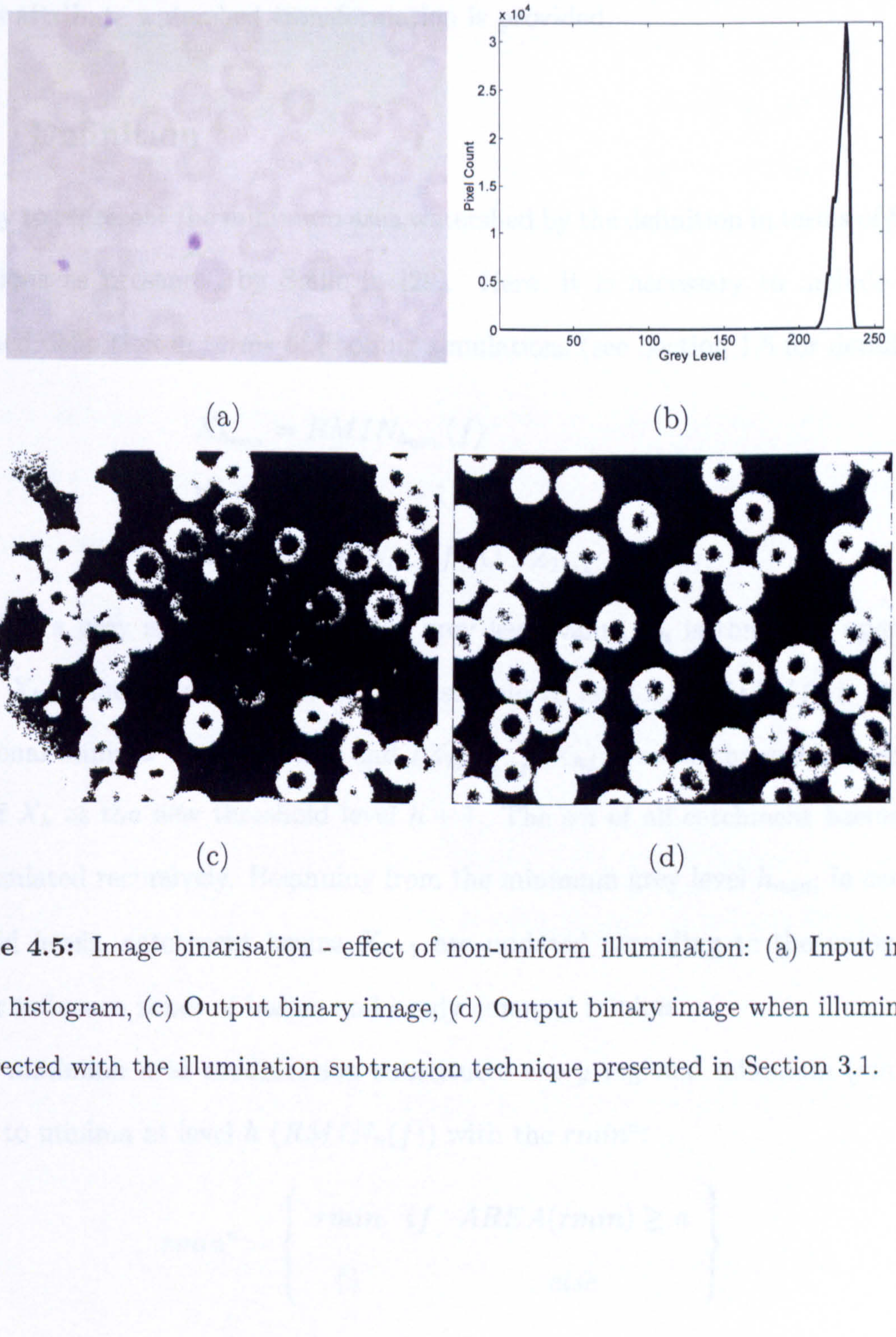


Figure 4.5: Image binarisation - effect of non-uniform illumination: (a) Input image, (b) its histogram, (c) Output binary image, (d) Output binary image when illumination is corrected with the illumination subtraction technique presented in Section 3.1.

are above a given threshold. The same concept is studied in [93] with a progressive merging strategy called an "attribute-based absorptions" (ABA) method. In [88] an area constrained watershed is briefly mentioned to overcome the over-segmentation problem. Here, the modification is formulated, an efficient algorithm is developed, the relation to the morphological area closing operator is discussed, and a generalisation towards attribute watershed transformation is provided.

4.4.1 Definition

It is easy to represent the minimum area watershed by the definition in terms of flooding simulations as presented by Soille in [29]. Here, it is necessary to include Soille's watershed definition in terms of flooding simulations (see Section 1.8 for details):

$$\begin{aligned} & X_{h_{min}} = RMIN_{h_{min}}(f) \\ & \forall h \in [h_{min}, h_{max} - 1], \\ & X_{h+1} = RMIN_{h+1}(f) \cup IZ_{T_{t \leq h+1(f)}}(X_h) \end{aligned} \tag{4.6}$$

where f is a grey scale image, h is the grey level value, T_h is threshold operator at level h, X_h is the set of the catchment basins up to level h, $RMIN_h(f)$ is the set of *all* regional minima at the level h, and $IZ_{T_{t \leq h+1(f)}}(X_h)$ denotes the geodesic influence zones of X_h at the new threshold level $h+1$. The set of all catchment basins $X_{h_{max}}$ are formulated recursively. Beginning from the minimum grey level h_{min}, in successive threshold levels, catchment basins X_{h+1} are updated according to the union of the geodesic influence zones of basins and newly emerged minima.

The minimum area modification substitutes *every* regional minimum (*rmin*) belonging to minima at level h ($RMIN_h(f)$) with the $rmin^a$:

$$rmin^a = \left\{ \begin{array}{cc} rmin & if \quad AREA(rmin) \geq a \\ \{\} & else \end{array} \right\} \tag{4.7}$$

4.4.2 Algorithm

The implementation is realised from Soille's watershed transformation algorithm [29] by modifying the last section about the new minima.

The three boxed lines (i.e. 5, 11, and 12) in Algorithm 1 show the modification for the minimum area watersheds. In the first line a simple counter (*A_counter*) is initialized for measuring area and an auxiliary queue is initialized by cleaning previous entries (*AQ_clean*). In the second boxed line, the pixel added to the FIFO is also added to the auxiliary queue for later use and the area counter is increased. When the process reaches the third boxed line, the pixels belonging to the current label are all registered in the auxiliary queue and the area is stored in the counter. Comparing the counter with the threshold (*A_th*), the algorithm resets the labels to the *MASK* if the area is not sufficient. Resetting the pixel labels to the *MASK* passes them to the next grey level $(h+1)$. That they will be regarded as being the same as pixels of level $h+1$. This process is iteratively repeated at successive grey levels until *A_th* is met.

To summarize, a region is not recognised as distinct until it has achieved a certain size (area). In order to ensure this, an area measuring step and a control statement is added to the Soille's algorithm. If the area of the newly emerged minimum is greater than the specified threshold it is labelled, else it is left unlabelled and tested again at the next grey level. The output of the original watershed algorithm is the same if the threshold is chosen as $A\_th = 1$. The MAWT algorithm reduces the number of calculated regions as the threshold is increased.

4.4.3 Relation to Area Closing and Extension to Other Attributes

As proved in [42], the relation between area opening and regional maxima can be stated as:

$$\gamma_\lambda^a(f) \quad \rightarrow \quad AREA(RMAX(f)) \geq \lambda$$

Algorithm 1 MAWT algorithm in pseudo-code

fi: (GREY SCALE INPUT IMAGE),fo: (OUTPUT IMAGE), A_th: (AREA THRESHOLD)

h: (GREY LEVEL VALUE), $A\_counter$: (AREA COUNTER)

$fifo\_add$: (ADD PIXEL TO TOP OF FIFO QUEUE),

$fifo\_empty$: (RETURN 1 IF FIFO EMPTY),

$fifo\_retrieve$: (RETRIEVE LAST PIXEL AND DELETE FROM QUEUE)

AQ_clean: (EMPTY AUXILIARY QUEUE), AQ_add: (ADD PIXEL TO AUXILIARY QUEUE),

AQ_remask: (RELABEL ALL PIXELS IN AUXILIARY QUEUE AS $MASK$)

...

{check for new minima}

1. $\forall$ pixel p such that $f_i(p) = h$ {
2. if $f_o(p) = MASK$ {
3. $current\_label+ = 1$;
4. fifo_add(p); $f_o(p) \leftarrow current\_label$;
5. $A\_counter \leftarrow 1$; $AQ\_clean()$; $AQ\_add(p)$;
6. while $fifo\_empty() = false$ {
7. $p' \leftarrow fifo\_retrieve()$;
8. $\forall$ pixel $p'' \in N_G(p')$ {
9. if $f_o(p'') = MASK$ {
10. $fifo\_add(p'')$; $f_o(p'') \leftarrow current\_label$;
11. $AQ\_add(p'')$; $A\_counter+ = 1$; }}}}

{now check the area}

12. if $A\_counter < A\_th$ $AQ\_remask()$; }

and by duality the relation between area closing and regional minimum is such:

$\phi^a_\lambda(f) \rightarrow AREA(RMIN(f)) \geq \lambda$

The area closing (ϕ^a_λ) operation on a grey level image (f) ensures the area of every regional minimum in $RMIN(f)$ is greater than or equal the closing parameter λ. Therefore every regional minimum $rmin$ will be equal to $rmin^a$ as in (4.7). This suggests the area closing (ϕ^a_λ) followed by the watershed transform is equivalent to the MAWT with area threshold (A_th) equal to λ. This is valid for all values including $\lambda = 1$.

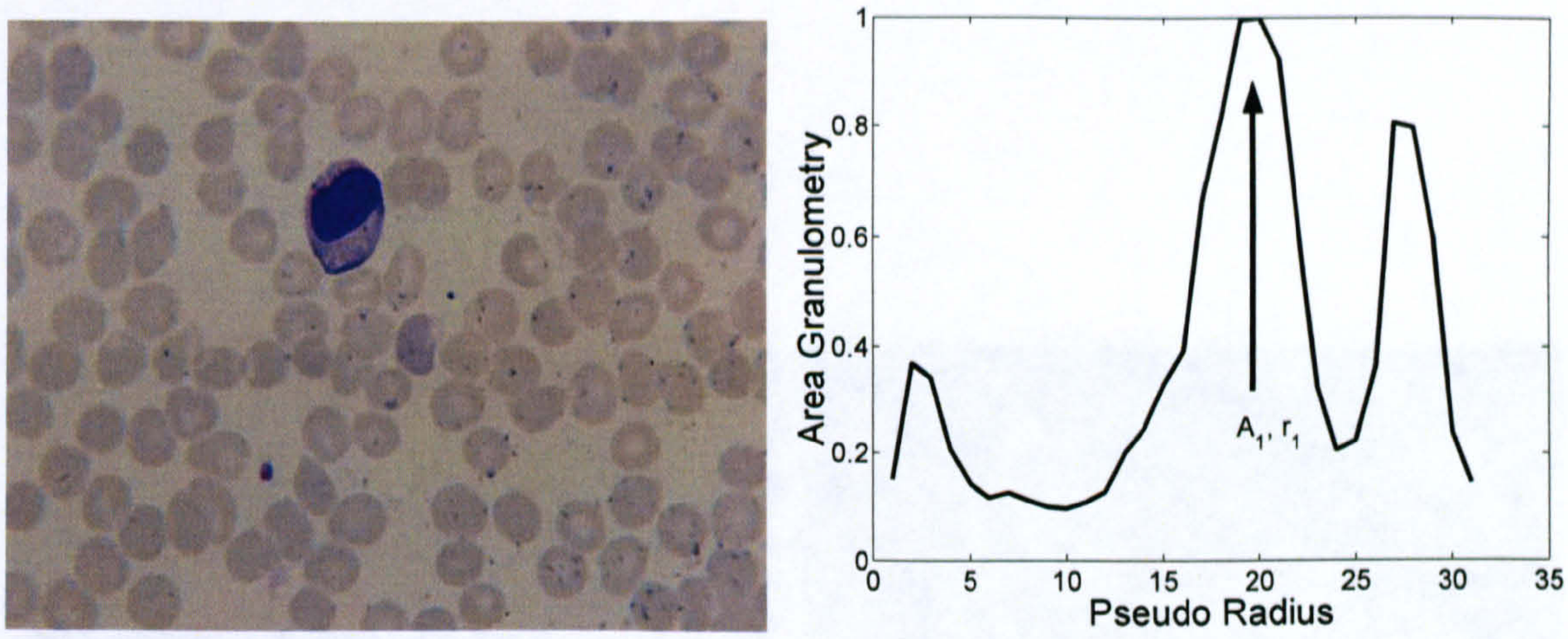
As stated earlier, MAWT checks the area of every regional minimum against the parameter λ and passes them to the next grey level ($h + 1$) if the area is smaller. Passing to next grey level causes the pixel to be processed as it is in $h + 1$. This shows that the MAWT algorithm performs as an embedded area closing ϕ^a_λ operation in the watershed transformation.

The connection between the area closing and the MAWT leads us to think about more attributes which are already used in attribute closings or openings [75]. The increasing attributes such as depth, volume or area of the minimum enclosing rectangle can be useful [93], and can be implemented by applying direct substitutions in Algorithm 1. A generalised attribute watershed transform covering non-increasing attribute functions (ATT) such as circularity or compactness can be obtained by introducing a generalised attribute criterion ($Criterion$) which is checking a given set of attributes (att) instead of the MAWT comparison of the area against a threshold:

$$rmin^c = \left\{ \begin{array}{cc} rmin & if \quad Criterion(ATT(rmin), att) = TRUE \\ \{\} & else \end{array} \right\} \tag{4.8}$$

4.4.4 Application

An initial segmentation (S_1) is calculated on the morphological image gradient using the MAWT algorithm. Since we can calculate the area distribution and A_1 from the



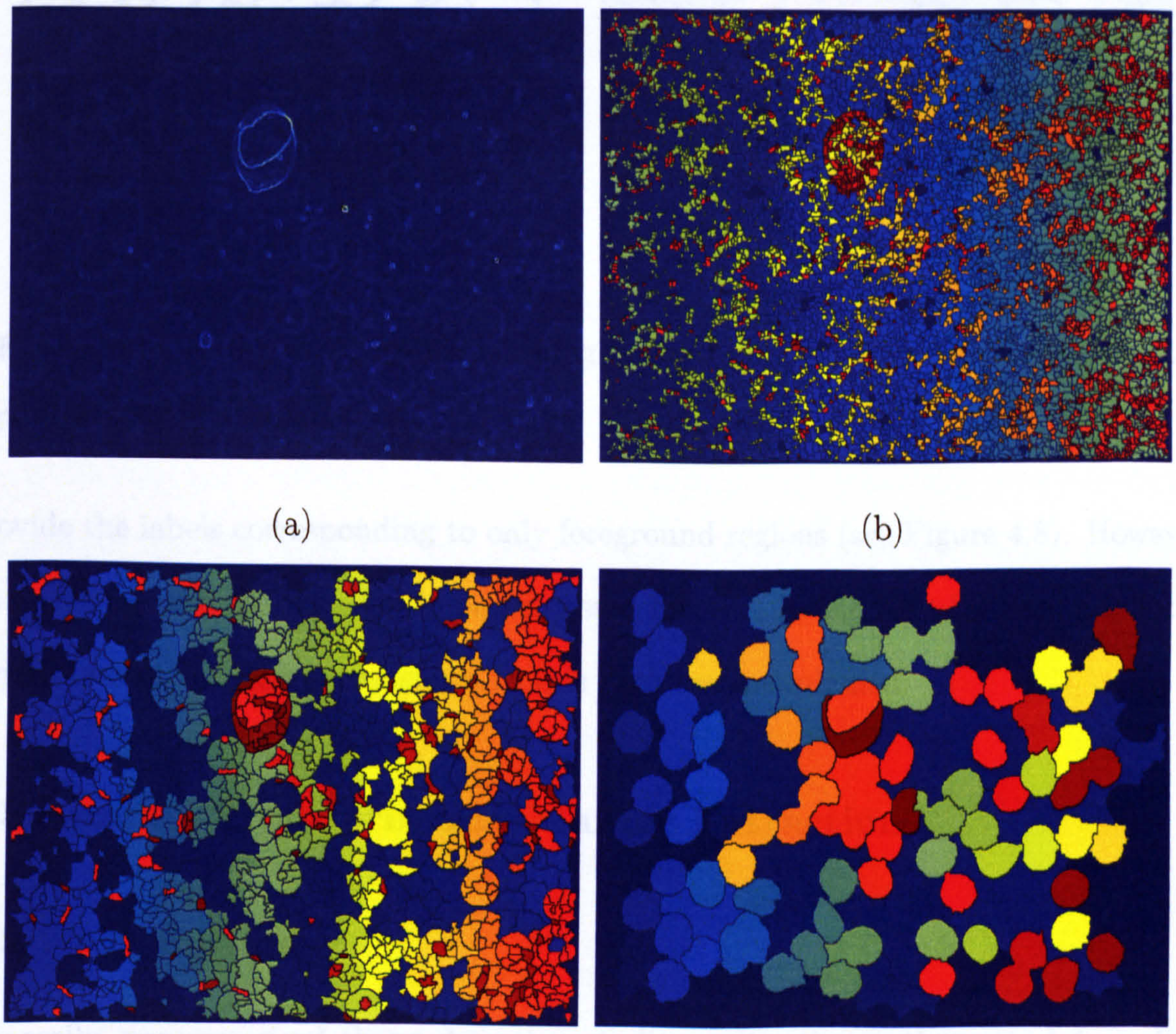
(a) (b)

Figure 4.6: Initial segmentation using MAWT - area granulometry calculation: (a) Input image (b) the area granulometry

area granulometry, it is possible to use it as a threshold for the regional minima area. Figure 4.6 shows an input thin blood film image and the calculated granulometry that identifies peak area and the corresponding pseudo-radius.

However, the blood cells are not flat, detached regions. Instead of using A_1, using a fraction (e.g. $\frac{A_1}{2}$) of it can be sufficient for the initial segmentation. In Figure 4.7(a) the morphological gradient of the input image (Figure 4.6(a)) is shown. The labelled output images are presented in Figure 4.7(b), 4.7(c), and 4.7(d), respectively for the original, MAWT with $A\_th = 10$, and $A\_th = \frac{A_1}{2}$. Figure 4.7(d) represents the labelled initial segmentation image (S_1) which will be used in the next steps of the method. The S_1 is not a binary or a grey level image: all pixels on the watershed lines have label zero and all other regions have unique labels greater than zero.

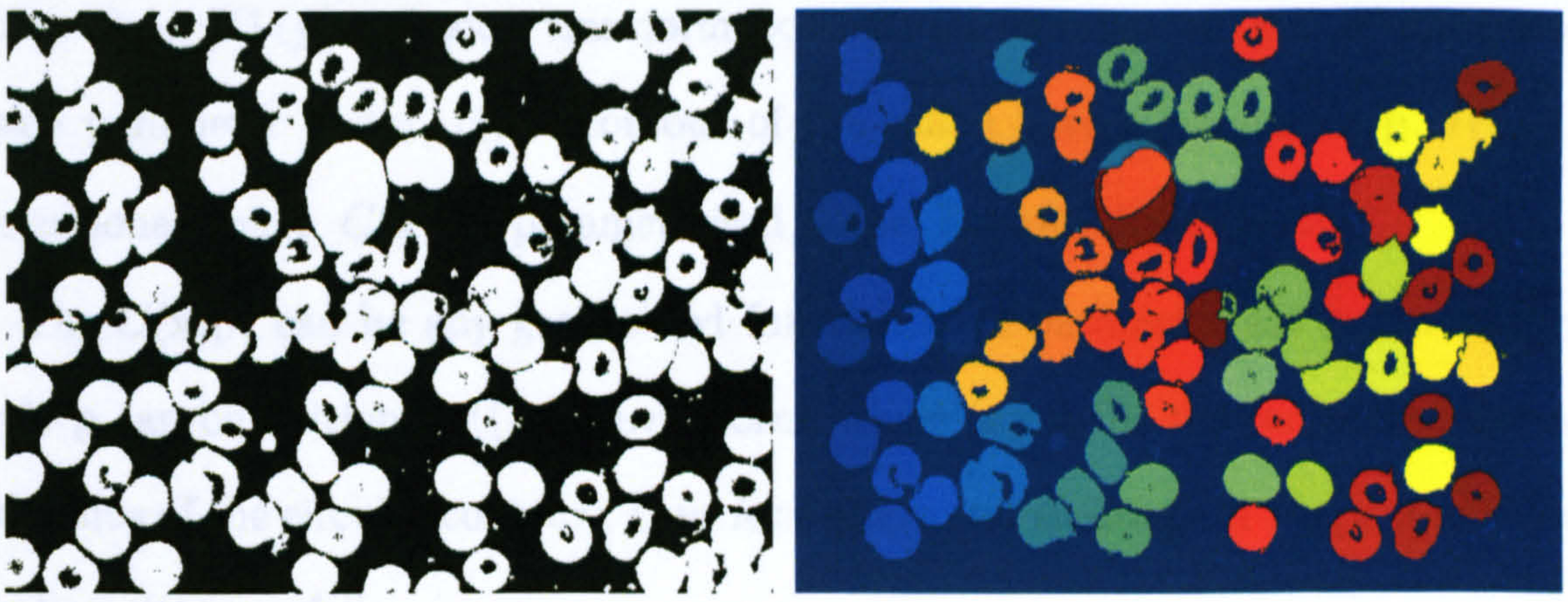
Although MAWT reduces the number of regions and calculates useful boundaries, it is not sufficient to complete the segmentation. As can be seen in Figure 4.7, the output also contains background regions and under-segmented regions. These problems must be solved with further processing. Background regions are simple to eliminate by using the binary foreground mask that is obtained with the image binarisation method explained in Section 4.3. Simply the pixel-wise multiplication $S_1. * B$ can



(a) (b)

(c) (d)

Figure 4.7: Initial segmentation using MAWT - comparison with original watershed algorithm: (a) Image gradient (b) watershed labels with the original algorithm and (c) watershed labels with the MAWT algorithm $A\_th = 10$ applied to gradient, (d) watershed labels (S_1) with the MAWT algorithm $A\_th = \frac{A_1}{2}$ applied to gradient. Note both foreground and background regions are labelled.



(a) (b)

Figure 4.8: Background elimination using binary mask: (a) foreground mask, (b) its pixel-wise multiplication by S_1.

provide the labels corresponding to only foreground regions (see Figure 4.8). However, under-segmented regions have to be processed further. In the next Section, a possible application of the Radon transform to this problem is explained.

4.5 Radon Transform and Marker Extraction

4.5.1 Definition

Generally, parameterised shape detection studies utilise the mathematical formalism of the Hough transform instead of the Radon transform. However, a recent study has shown that the two methods are equivalent [96]. The difference is simply in their computational interpretations. If the data is not sparse, the Hough transform is computationally more expensive than the Radon transform. However, there are several applications and realisations of the Hough transformation in the literature [119–123] while far fewer for Radon transformation [124–126]. The contour integral description that exists in the Radon transformation enables the implementation via convolution [96]:

$$R_{c(\mathbf{p})}\{I\}(\mathbf{p}) = \int_{c(\mathbf{p})} I(\mathbf{x})\, d\mathbf{x} = \int_{\Re^D} \delta(C(\mathbf{x};\mathbf{p}))\, I(\mathbf{x}) d\mathbf{x}, \tag{4.9}$$

where $R_{c(\mathbf{p})}\{I\}(\mathbf{p})$: Radon transform, $\mathbf{x}$: spatial coordinates, $\mathbf{p}$: N-dimensional shape parameter vector, $c(\mathbf{p})$: contour of the shape having parameter $\mathbf{p}$, $I(x)$: D dimensional image, $C(\mathbf{x};\mathbf{p})$: parameterised shape, δ: delta-dirac function.

The $C(\mathbf{x};\mathbf{p})$ can be any generalised function with parameter set $\mathbf{p}$. In case of a circle $\mathbf{p}$ can be written in $((x_0, y_0), r_0)$ form, where (x_0, y_0) is the centre and the r_0 is the radius of the circular contour. Thus for a circle and an image of $I(x, y)$ the formula can be written as follows:

$$R_{c((x_0,y_0),r_0)}\{I\} = \int_{\Re^2} \delta((x - x_0)^2 + (y - y_0)^2, r_0^2)\, I(x, y)dx\, dy \tag{4.10}$$

If the transformation is to produce a response on every pixel (every possible (x_0, y_0)), the above continuous integral can be transformed to a convolution with a shift-invariant circle kernel having radius r_0. Excluding (x_0, y_0) from our parameter vector, $\mathbf{p}$ reduces to variable r. Thus, we can rewrite the formula as:

$$R_{(r)}\{I\} = I * K_r, \tag{4.11}$$

where $K_r = \delta(x^2 + y^2, r^2)$.

Generally the image I in (4.11) is a binary edge representation of the original image. The edge image (E_I) can be obtained by calculating the difference between S_1 and the erosion of S_1 (i.e. the erosion gradient) with structuring element B.

$$E_I = S_1 - \epsilon_B(S_1) \tag{4.12}$$

Since S_1 is labelled, the edge image E_I provides the region contours with unique labels. Label information can be useful if we incorporate it into our formula. However, there is no direct substitution between E_I and I for (4.11) such that: $R(r)\{I\} = E_I * K_r$. Instead, each labelled contour's R should be calculated separately:

$$L = \{l_1, l_2, l_3, ..., l_N\}, \quad R^L(r) = \bigcup_{l=1}^{N} \left\{E_I^{(l)} * K_r\right\} \tag{4.13}$$

where l_i denotes the ith label and $E_I^{(l)}$ connected edge pixels belonging to the ith label in the edge image.

To summarise, a labelled circle Radon transformation is derived which can be calculated on the labelled contour representation. Instead of searching for resolvable circles and detecting the correct radius, the main interest here is to find centres of the objects with radius closer to the value r_1 which was estimated by the area granulometry. In case of more than one radius, we can calculate a circularity map in labels by calculating *argmax* of the R^L for the individual radius in the set $\mathbf{r} = r_1, r_2, \ldots, r_k$.

$$R^L_{\mathbf{r}} = argmax_{\mathbf{r}}(R^L(\mathbf{r})) \tag{4.14}$$

The above calculation provides a 2-d mapping for the circular region centres and reduces calculations of different r to a single concern.

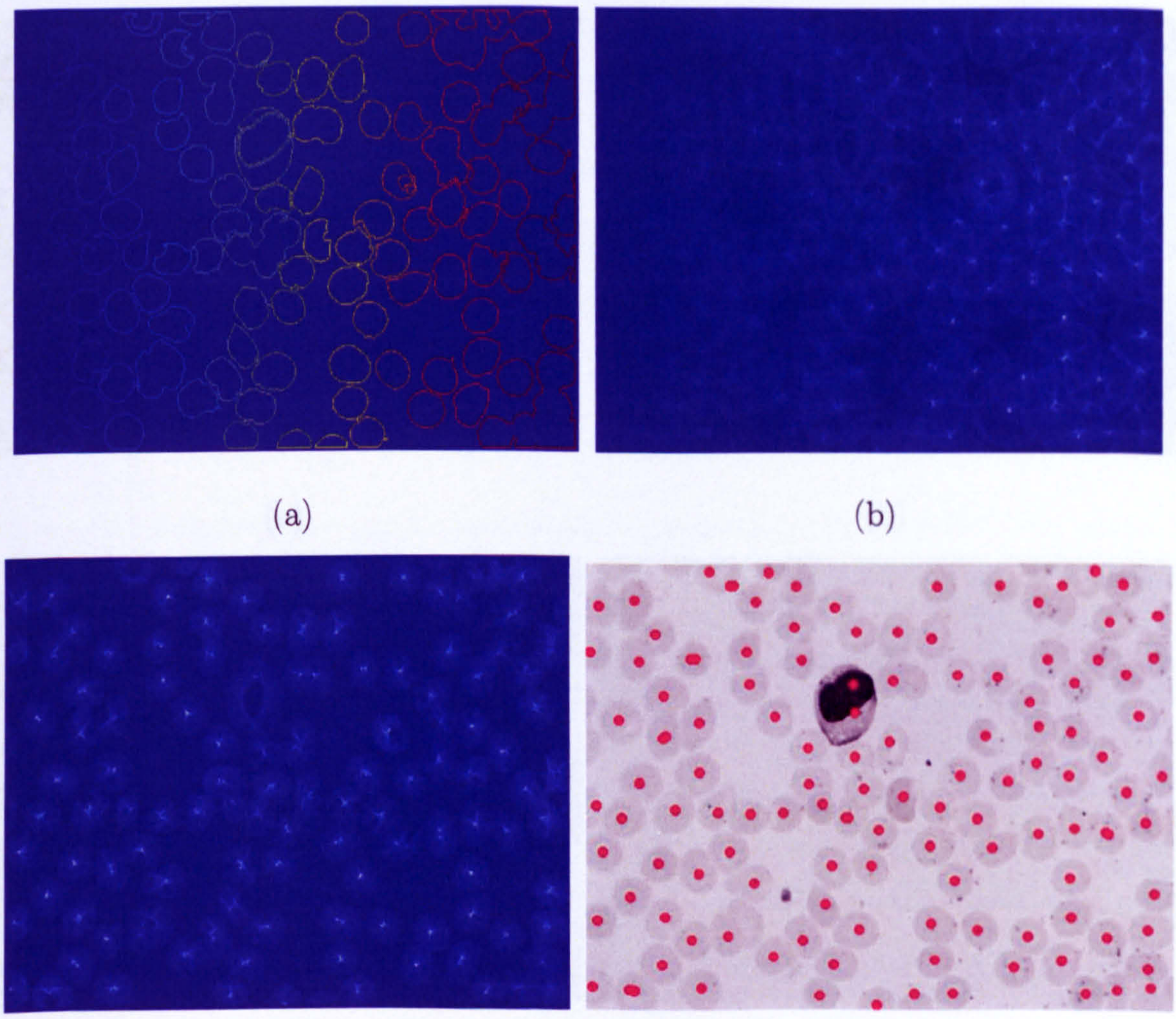
4.5.2 Application

The labelled circle Radon transformation is used to calculate the centres of labelled boundary regions obtained from the initial segmentation of the foreground objects. Figure 4.9(a) shows the labelled edge image E_I computed according to (4.12). The circularity map ($R^L_{r_1}$) calculated for r_1 (4.13) on the E_I in Figure 4.9(b). An unlabelled calculation of the transformation (R) (by direct convolution with thresholded E_I, i.e. $E_I > 0$) is presented for comparison in Figure 4.9(c). The latter is noisier due to the effect of the adjacent region boundaries.

After $R^L_{r_1}$ is calculated, markers (M) can be extracted by searching for the regional maxima. Regional maxima are calculated with a structuring element defined as a disk of radius $\frac{r_1}{4}$, then dilated to unite very close points (Figure 4.9(d)).

4.6 Under/Over-Segmented Regions

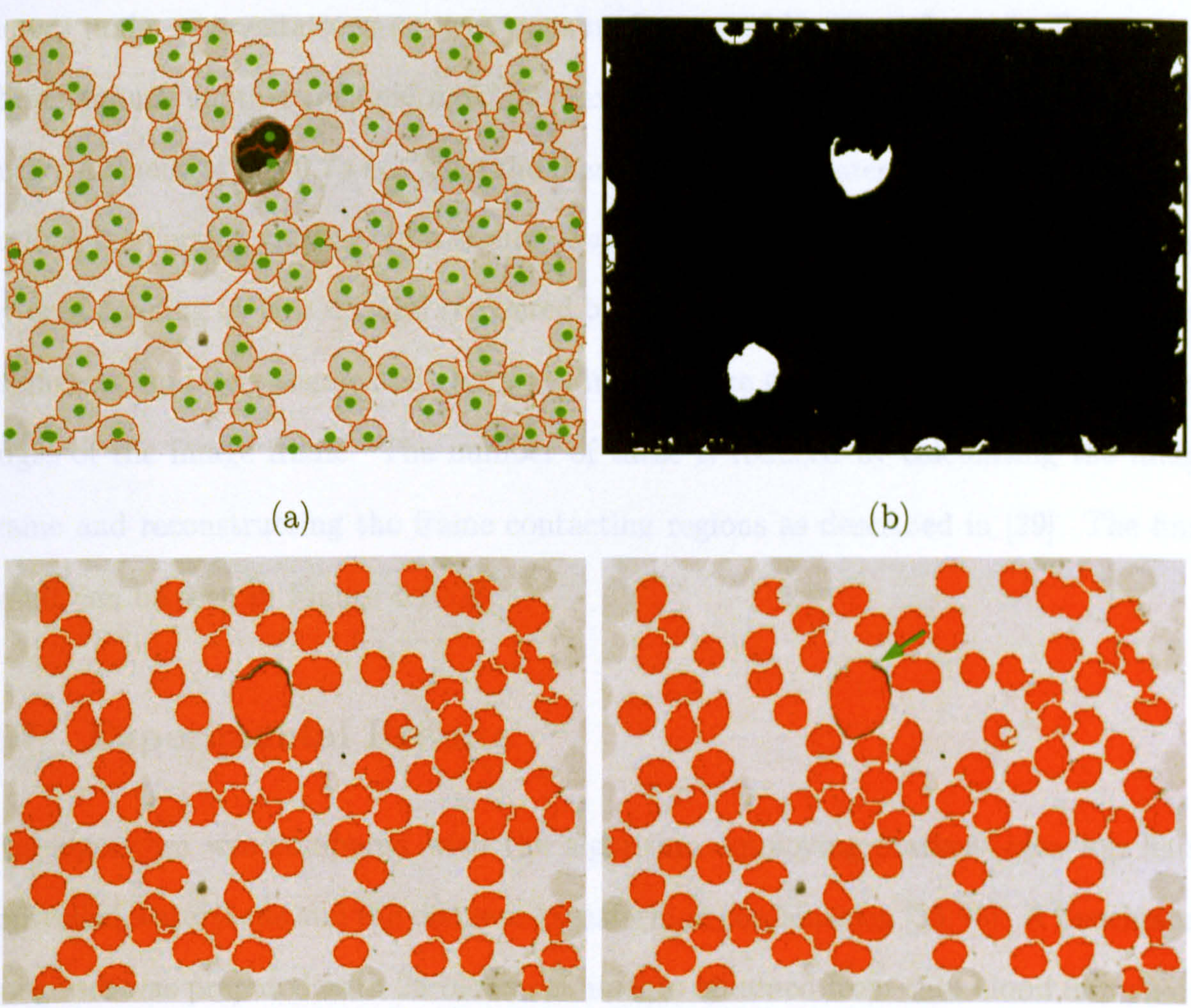
To finalise the segmentation, we apply the marker controlled watershed transformation using the markers calculated in the previous step M. For the input, morphological image gradient is used again (Figure 4.7(a)).



(a) (b)

(c) (d)

Figure 4.9: Radon transform based marker extraction and comparison: (a) The labelled edge image (E_I) for S_1 (b) Unlabelled Radon Transform by direct convolution for thresholded E_I (i.e $E_I > 0$) with disk radius r_1, (c) labelled circle Radon transform $R^L_{r_1}$ on E_I, (d) Markers extracted from $R^L_{r_1}$ shown superimposed on the grey level image.



(a) (b)

(c) (d)

Figure 4.10: Watershed lines calculated with the extracted markers: (a) Watershed lines corresponding to markers, (b) over- and under-segmented regions, (c) final segmentation output (d) final segmentation output when the stained objects are marked beforehand.

In Figure 4.10(a) the markers and the watershed lines corresponding to these markers are shown. However, the marker extraction algorithm in the marker extraction step depends on the regional maxima, and the size of the structuring element used in calculation. Hence, it can still result in under-segmented or over-segmented regions. To detect under-segmentations an area opening is applied by size $1.5 * \pi r_1^2$. To separate these regions, another regional maximum extraction process is applied by a disk structuring element of $r = 0.7 * r_1$. Then the marker controlled watershed is employed again to find the correct regions. In a similar way, the over-segmented regions are detected by area opening by size $\frac{A_1}{2}$, and recovered by dilation. Figure 4.10(b) shows the under-segmented and over-segmented regions. Finally, there are some cells connected to the edges of the image frame. The number of these is reduced by calculating the image frame and reconstructing the frame contacting regions as described in [29]. The final result can be seen in Figure 4.10(c).

4.7 Experimental Results

The algorithm was compared with the algorithm employing marker (area top hats) controlled watersheds and the distance transform in combination [35,95]. A benchmark image set was prepared with 25 blood cell images obtained from thin blood film slides. The set contained 11 new images from UoW microscope imaging setup in addition to 14 images that were supplied by NIMR which were used in the previous studies [35,95]. The images were chosen from fields which are not very concentrated and contained 2450 resolvable cells including RBCs, WBCs and RBCs infected by malaria parasites. The cells were manually marked to provide the ground truth data as explained in Section 3.2.2. The comparison was done by evaluating the "locating" and the "counting" performances.

The missed and under-segmentation rates were calculated directly by comparing labelled cell locations to manual markers. The over and redundant segmentation rates

were calculated from all labelled regions which did not coincide with a manually marked cell (Figure 4.11(a)) (Algorithm 2).

Algorithm 2 Segmentation performance evaluation algorithm

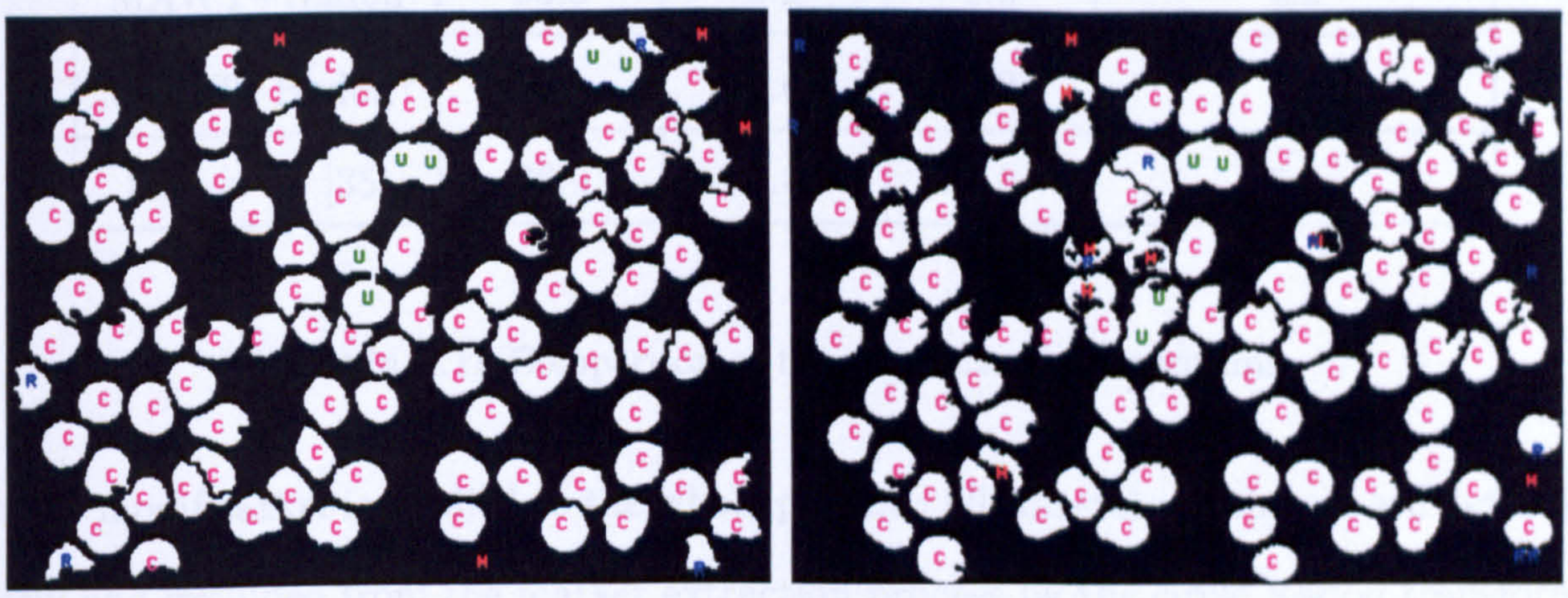
P : SET OF THE MANUALLY MARKED SET OF POINTS B : THE LABELLED OUTPUT IMAGE OF THE SEGMENTATION N_m : NUMBER OF MISSED CELLS N_c : NUMBER OF LOCATED CELLS N_u : NUMBER OF UNDER-SEGMENTED CELLS N_{or}: NUMBER OF OVER-SEGMENTED CELLS AND REDUNDANT SEGMENTS

$Init(counters)$

1. $\forall p \in P$ {
2. if $(B(p) = 0)$
3. $N_m \leftarrow N_m + 1$;
4. elseif $(\forall p' \in P, p' \neq p, B(p') \neq B(p))$
5. $N_c \leftarrow N_c + 1$;
6. elseif $(\exists p' \in P, p' \neq p, B(p') = B(p))$
7. $N_u \leftarrow N_u + 1$; }
8. $N_{or} \leftarrow$ number of unclassified remaining labels in B

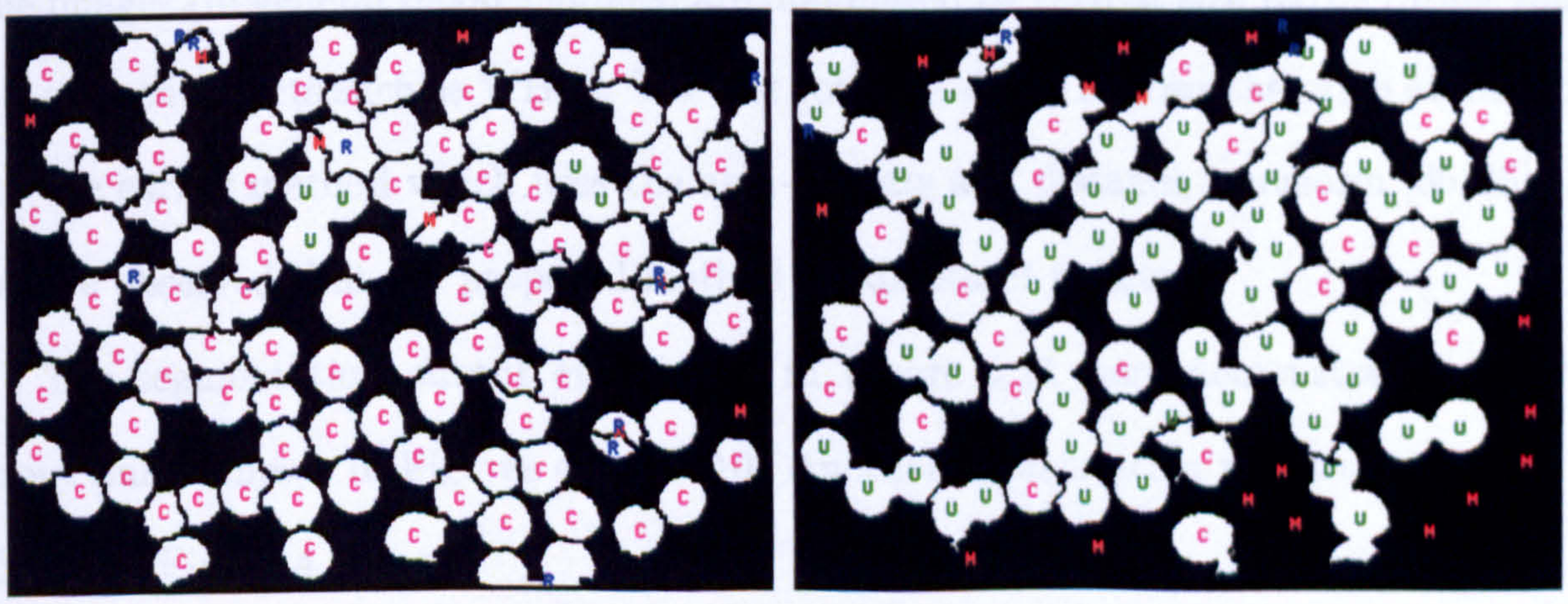
For every manually marked point p, a corresponding label is checked in the output. If the corresponding label is 0 (background), the segmentation has *missed* the cell marked by p. Else if the corresponding label is not 0, and no other $p' \in P$ which has the same corresponding label, the segmentation is *correct* for p. Else if there is at least one other $p' \in P$ has the same corresponding label, p is *under*-segmented. After all $p \in P$ have been classified, the remaining labels in the output are counted as *over&redundant* segmentations.

Table 4.1 shows evaluated results in four categories. The evaluated segmentation performance rates of the proposed method are shown in the first row. Additionally a



(a) (b)

Figure 4.11: Segmentation Evaluation example: (a) Evaluation marks: correct (pink 'C'), under (green 'U'), missed (red 'M'), over and redundant (blue 'R') segmentations. (b) The same evaluation for the result obtained by Rao's method [35,95].



(a) (b)

Figure 4.12: Segmentation Evaluation example: (a) Evaluated output for an input image (Figure 6.2(b)) of the proposed method, (b) Rao's method [35,95].

Method	correct %	missed %	under %	over&redundant %
MAWT+Radon	90.8	6.2	2.9	4.2
MAWT+Radon*	91.1	5.6	3.3	3.8
Rao *et al.* [35]	86.9	6.4	6.6	5.5

Table 4.1: Comparison of the segmentation algorithms

minor modification is applied to the whole process: all the stained objects are detected early and excluded from the marker extraction process by the circle Radon transform. This protects the WBCs and other stained objects from being over-segmented. Figure 4.10(d) shows the output of the segmentation process when the stained objects are marked and excluded from the marker extraction process. The results demonstrated in the second row of Table 4.1 were obtained by this modified algorithm. It can be observed the resulting rates were improved when the stained objects were marked prior to segmentation (the second row): the correct segmentation rate increased; the over-segmentation rate decreased. The difference is significant if we consider the fact that the images (in general blood samples) are dominated by regular size RBCs rather than big stained objects such as WBCs. The last row in Table 4.1 shows the same evaluation for the earlier method which uses the area-tophats and distance transform [35, 95]. It can be observed that the proposed method yields better results.

The segmentation output evaluations in Figure 4.11 were demonstrated for the input image shown in Figures 4.6 to 4.10. Another segmentation result comparison is shown in Figure 4.12.

Finally, to demonstrate the effect of the cell concentration of the image to the segmentation, two input images and corresponding segmentation results of the proposed method and the earlier method are shown in Figure 4.13. It can be observed that there are many under-segmented regions; it is hard to suggest that the segmented regions correspond to the actual cells in both of the methods. As a consequence, for the first

input image, the infected cell located in the centre is incorrectly merged to a bigger group (Figures 4.13(c) and 4.13(e)). However, for the second image, the infected cell is successfully found because it is isolated on the background (Figures 4.13(d) and 4.13(f)).

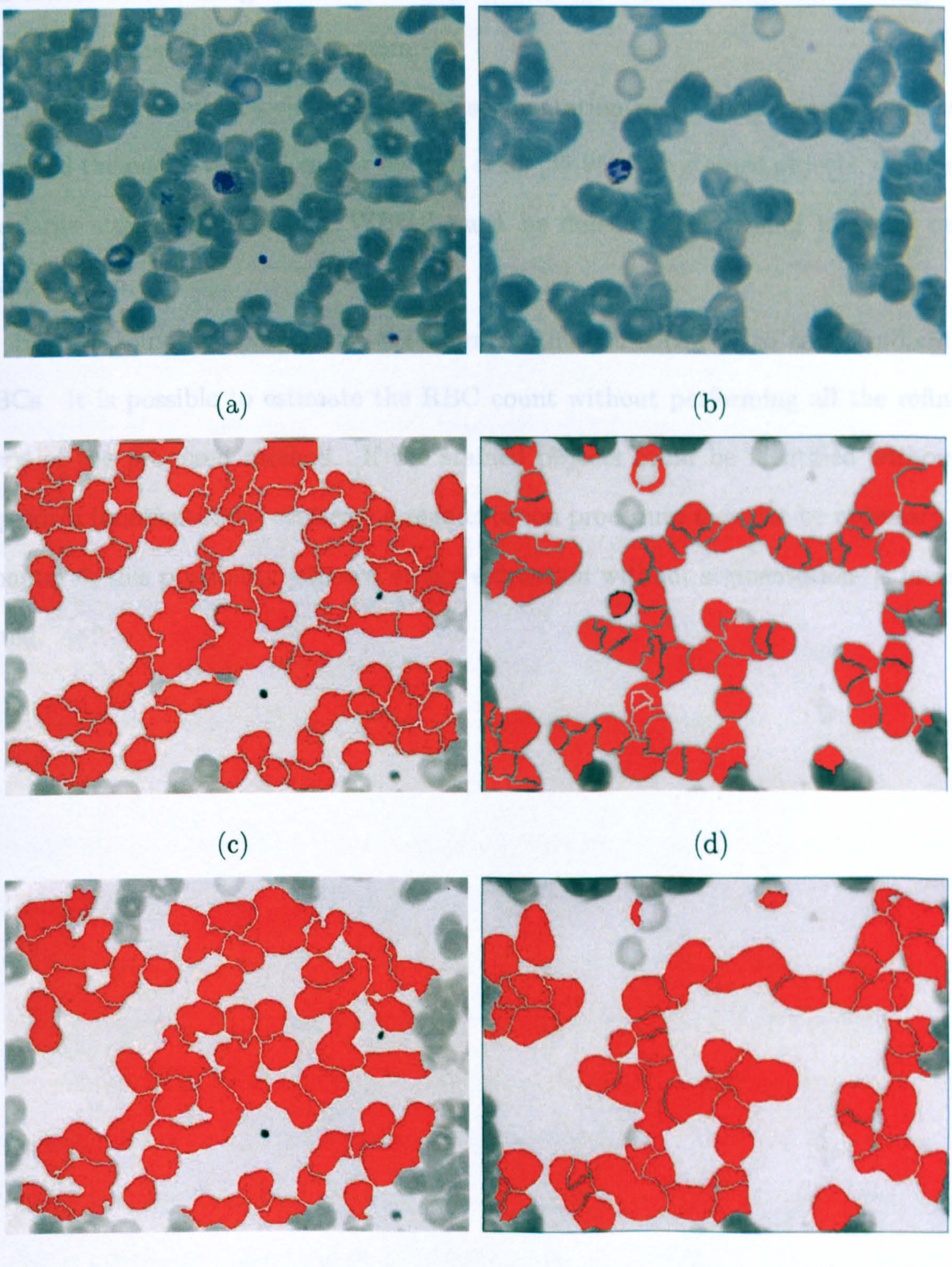
4.8 Conclusions and Discussions

In this study, a new thin blood film image segmentation method has been presented. The method defines a new MM tool -*minimum watershed transformation* (MAWT)- and applies it to perform an initial segmentation. This modified watershed transformation introduces a solution for the over-segmentation problem associated with the classical watershed algorithm [47]. Therefore, it eliminates the necessity for marker calculation prior to watershed segmentation. It is possible to extend the MAWT to cover other attributes. This is identified as future work.

If applied to blood cell images, the segmentation result with minimum area watershed transform using the average cell area estimate contains under-segmented regions. To reduce these regions, a circular object centre locater algorithm has been derived from Radon transform [96]. The final result is produced by applying the marker controlled watershed transform [29] with markers obtained from the circle Radon transform. In order to prevent over-segmentation of stained objects, they were extracted first and excluded from the marker extraction process.

The proposed method was compared to the method developed by Rao *et al.* [35,95] which is the best earlier method. The evaluations of the cell locating and counting performances demonstrated that the new algorithm produces better results.

However, both of the algorithms are not applicable to concentrated fields of thin blood film slides. As explained in the introduction the aim of the segmentation process demonstrated here is to define the cell regions for the later processes of the CDM. Thus, the segmentation task introduces errors for the further analysis unless the images



(a) (b)

(c) (d)

(e) (f)

Figure 4.13: Segmentation in concentrated fields: (a-b) input images, (c-d) segmentation output of the proposed method, (e-f) Rao's method [35, 95].

are selected from only less concentrated fields. It is possible to evaluate an image's cell concentration using the method in [26]; however this will degrade the diagnosis capability of the overall CDM system.

Another problem associated with the segmentation method that is demonstrated here and the one developed earlier by Rao *et al.* [35,95]: the stained objects, especially the large stained objects (e.g. WBCs) must be detected beforehand to avoid over-segmentation.

However, another purpose of the segmentation for CDM is to count individual RBCs. It is possible to estimate the RBC count without performing all the refining steps of the proposed method. If the stained objects could be identified without a top-down (segmentation) strategy, a segmentation procedure may not be necessary. In Chapter 6, this possibility -stained object extraction without segmentation- is investigated.

Chapter 5

Colour Normalisation

5.1 Introduction

Microscopy diagnosis of malaria infection requires recognition of *Plasmodium* on stained blood films. The different Plasmodium species are distinguishable from each other and regular blood components or artefacts by their characteristic shapes (morphology) and colour properties. For example, the Hemozoine pigment for *P. Falciparum* which is an identifier of mature stage parasites is described by manual diagnosis experts as a brownish or yellowish region inside the parasite: light blue-purple cytoplasm covering dark blue nuclei with a species specific shape. This type of colour definition can be definitive enough for the human visual system; however, it may be ambiguous for computer vision. An intuitive approach to remove the ambiguity could be to define a colour space such as RGB and define the term *brownish* with a subspace defined by boundaries. However, some questions would immediately arise such as "what is the reference to the *brownish*" or more technically "where should the origin (0,0,0) of the colour space be placed?". Since it is practically impossible to define an absolute universal origin, in computer vision this question is answered by calibration: obtain the imaging equipment for which the colour *brownish* is defined, find its response to a defined reference (e.g. *white* patch) and transform the origin of the imaging equipment's colour space according to this reference. Then the *brownish* colour will correspond to a specific set of coordinates in this transformed (according to the reference) colour space.

An image acquired from a stained blood sample (thick or thin) using a conventional light microscope can have several conditions which may affect the observed colours of the cells, plasma (background), and stained objects. These conditions may be due to the microscope components such as: the different colour characteristics of the light source, intensity adjustments, or colour filters. They may be due to the use of different

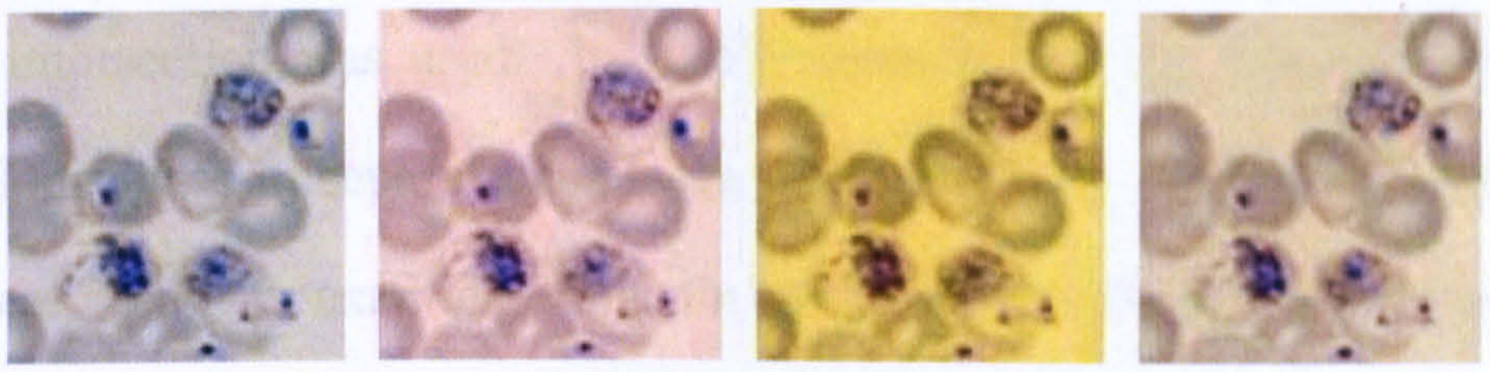


Figure 5.1: Example images of the same thin blood film (rodent) field with different chromatic filters.

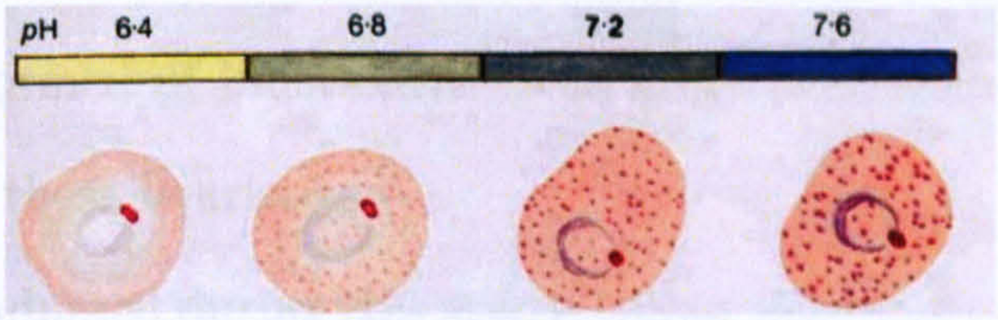


Figure 5.2: Effects of different staining (i.e. PH) on an infected cell, reproduced by permission [9].

cameras or different settings in the same camera: exposure, aperture diagram, or white balance settings. The effects of these conditions may be reduced by calibration if a reference colour chart is possible. However, colour differences can also be related to the differences in the slide preparation process (i.e. the object itself): different stain concentrations, different staining durations, or non-uniform staining. Thus, a standardisation for the blood film preparation itself is also required. In Figure 5.1 different images which are acquired from the same field with different camera white balance settings can be observed. An illustration in Figure 5.2 from the WHO Basic Malaria Microscopy [9] demonstrates the effects of the different PH (i.e. acidity level) values (of the Giemsa-stain) on a blood cell occupied by a parasite. It can be observed that PH of the stain can affect significantly the appearance.

The earlier research on Computerised Diagnosis of Malaria (CDM) did not utilise all the colour information available on different channels. In Di Ruberto's research [34,127], the RGB images were transformed to the HSV colour space and the individual channels were used in different steps of analysis, separately. For example, stained pixels were separately detected on Saturation and Hue channels of the converted image, and

then their intersection was used. In Rao's research [35] the images were processed in RGB colour space in separate channels, although some relations between different channels were used. For example, the stain colour bias in the images was estimated by the peak index differences between blue and red or green channels of the images; the grey level image was used for segmentation; the green channel histogram was found to be more useful to extract stained objects; the mature parasite pigment Hemozoine was sought in an image formed of blue and red (or green) channel differences. It can be difficult to generalise these heuristics.

In the research conducted during this study, colour information was extensively used in the analysis of parasites. The stained pixels were detected according to the RGB colour vectors; in the parasite detection and species/life stage recognition tasks, some colour based features were used in classifiers. As explained in Chapter 3, in this thesis, many images are used which are acquired from different slides, under different imaging conditions. These images are affected by nearly all the factors that are mentioned in preceding paragraphs. Hence, it was required to investigate the possibility of reducing these effects by using "colour normalisation". The colour normalisation research for the peripheral thin blood film images aimed to produce several significant outcomes: to remove the necessity of the pre-calibration of the imaging setup; to improve the robustness of the system for non-standard slide preparations; to provide the flexibility of exchanging images and data between different malaria diagnosis laboratories that might use the CDM system in the future. This kind of flexibility can provide the CDM system with the ability to update/learn from samples coming from different sources and redistribute them to different laboratories to maintain a uniform diagnosis capability.

In this thesis study, a method is developed to maintain the colour constancy for the images captured from various (unknown) sources. The method is very practical and exploits the special characteristics of the peripheral thin blood film images.

The main contribution introduced in this Chapter is a new colour normalisation method for thin peripheral blood films. This study was presented in 14th Signal Pro-

cessing and Communication Applications Conference (SIU-2006) and published in the proceedings [39].

5.2 The Colour Normalisation Method

The primary aim of colour constancy is to obtain the colour characteristics of the object independent of the illuminant and imaging sensor. In general, these two factors are studied separately: illumination change and sensor (camera) calibration. Sometimes the latter is not considered (as in this study) when the sensor is not present or its calibration is not possible.

There are some different models of illumination change in the literature [58]. The "diagonal model" is a simple and satisfactory model which assumes that there is a diagonal $3x3$ linear transformation matrix which maps ($\mathbf{p}^c = \mathbf{M}\mathbf{p}^u$) the RGB response under an unknown illuminant $\mathbf{p}^u = (r^u, g^u, b^u)$ to the RGB response under a canonical known illuminant $\mathbf{p}^c = (r^c, g^c, b^c)$. When the transformation matrix is assumed to be diagonal (5.1), its non-zero elements (m_{ii}) can be calculated by simple scaling p_i^c/p_i^u where $i \in \{r, g, b\}$.

$$\mathbf{M} = \begin{bmatrix} m_{rr} & 0 & 0 \\ 0 & m_{gg} & 0 \\ 0 & 0 & m_{bb} \end{bmatrix} \tag{5.1}$$

If the illumination is assumed to be uniform, using the diagonal model, an image of unknown illumination $\mathbf{I}^u$ with RGB colour vector can be simply transformed to a known illuminant space $\mathbf{I}^c$ by multiplying all the pixel values with the diagonal matrix ($\mathbf{I}^c = \mathbf{M}\,\mathbf{I}^u$). If the illumination is varying locally, each pixel ($\forall p \in I^u$) has to be scaled individually with ($\mathbf{I}^c(p) = \mathbf{M_p}\,\mathbf{I}^u(p)$).

The most simple colour constancy algorithm is based on the assumption that the average value of the scene is stable (e.g. grey) under the same illumination (i.e. grey world assumption). In other words, the grey world assumption suggests that the deviation of the average value of the image of the scene reflects the illumination change.

The different interpretations of this assumption can lead to different approaches. For example, the average value of the scene can be assumed to be actually *grey* which is a portion (e.g. half) of the maximum possible value of each channel. Thus, using the diagonal model, the illuminant factors can be calculated as in (5.2):

$$m_i = \frac{G_i}{\mu_i^u} \qquad (\mu_i^u = \frac{1}{N}\sum_N I_i^u) \qquad i = \{r, g, b\} \tag{5.2}$$

, where G_i is the constant (assumed) grey value and the μ_i^u is the mean of the channel I_i^u having N total pixels .

Alternatively, the *grey* value(s) can be defined with respect to the average values of the scene under a known illuminant. This is called *the database grey-world algorithm* [128, 129] and can be represented as in (5.3):

$$m_i = \frac{\mu_i^c}{\mu_i^u} \qquad i = \{r, g, b\} \tag{5.3}$$

Thus the illumination factors, the non-zero elements of the diagonal matrix, m_{rr}, m_{gg}, and m_{bb} are calculated by the ratios of the average values of the each channel of the reference (μ_i^c) to those of unknown (μ_i^u).

For ordinary images, the illumination estimation or colour normalisation based on the grey world assumption with either of the calculations yields poorer results compared with more sophisticated algorithms such as gamut mapping, retinex or colour by correlation [58, 128, 129]. Obviously, the weakness of the grey world algorithm arises from the fact that unconstrained physical scenes having limited members of objects do not reflect a universal grey value.

However, the images subject to this study are constrained: they contain two basic components (plasma and the rest which is primarily red blood cells) which can be separated by foreground-background segmentation. The plasma of the peripheral thin blood film image can be assumed to be a colourless transparent region. It is possible to define the constant grey values G_i (5.2) for colourless transparent (expected) pixel colour. Hence, an image of unknown illumination can be transformed using these

values as shown in (5.2). In fact, we can simply assign the colourless transparent pixel colour as the maximum possible value of a pixel (255) for all channels (each channel is commonly quantised to 256 discrete levels).

On the contrary, because of the effects of staining on the foreground objects the same assumption can not be made. In other words, we can not simply define or assume a universal average value for the cells since they may not be homogeneous due to stained components. However, it is possible to define the grey values according to a reference (i.e. the database grey world) and to transform the pixels for the cells (foreground objects) using (5.3). Thus, the reference which is assumed to taken under known illuminant imposes the constant grey values.

Therefore, the simple grey value assumption can be incorporated into the colour normalisation process of thin blood films. However, initially, input images must be separated into foreground and background regions. The separation (i.e. binarisation) can be performed using the method proposed by [35], which was described in Section 4.3. This method uses area morphology to estimate size of the cells and then extracts foreground objects and estimates individual histograms for the foreground and background regions. Then these histograms are used to obtain thresholds to perform a morphological double thresholding operation [29].

After separating the input ($\mathbf{I}_i^u$) channels ($i \in \{r, g, b\}$), foreground ($\mathbf{I_f}_{r,g,b}^u$) and background ($\mathbf{I_b}_i^u$) images are obtained and the proposed colour normalisation is performed as follows:

1. Calculate ($\mathbf{I_b}_i^u$) channel averages. Calculate $\mathbf{M^b}$: $m_i^b = \frac{255}{\mu^{\mathbf{I_b}_i^u}}$.

2. Transform the whole image: $\mathbf{I^1} = \mathbf{M^b}\ \mathbf{I^u}$ with equation (5.2)

3. Calculate $\mathbf{M^f}$: m_i^f using equation (5.3) with ($\mathbf{I_f}_i^1$) and the reference image foreground channels $\mathbf{I_f}_i^c$.

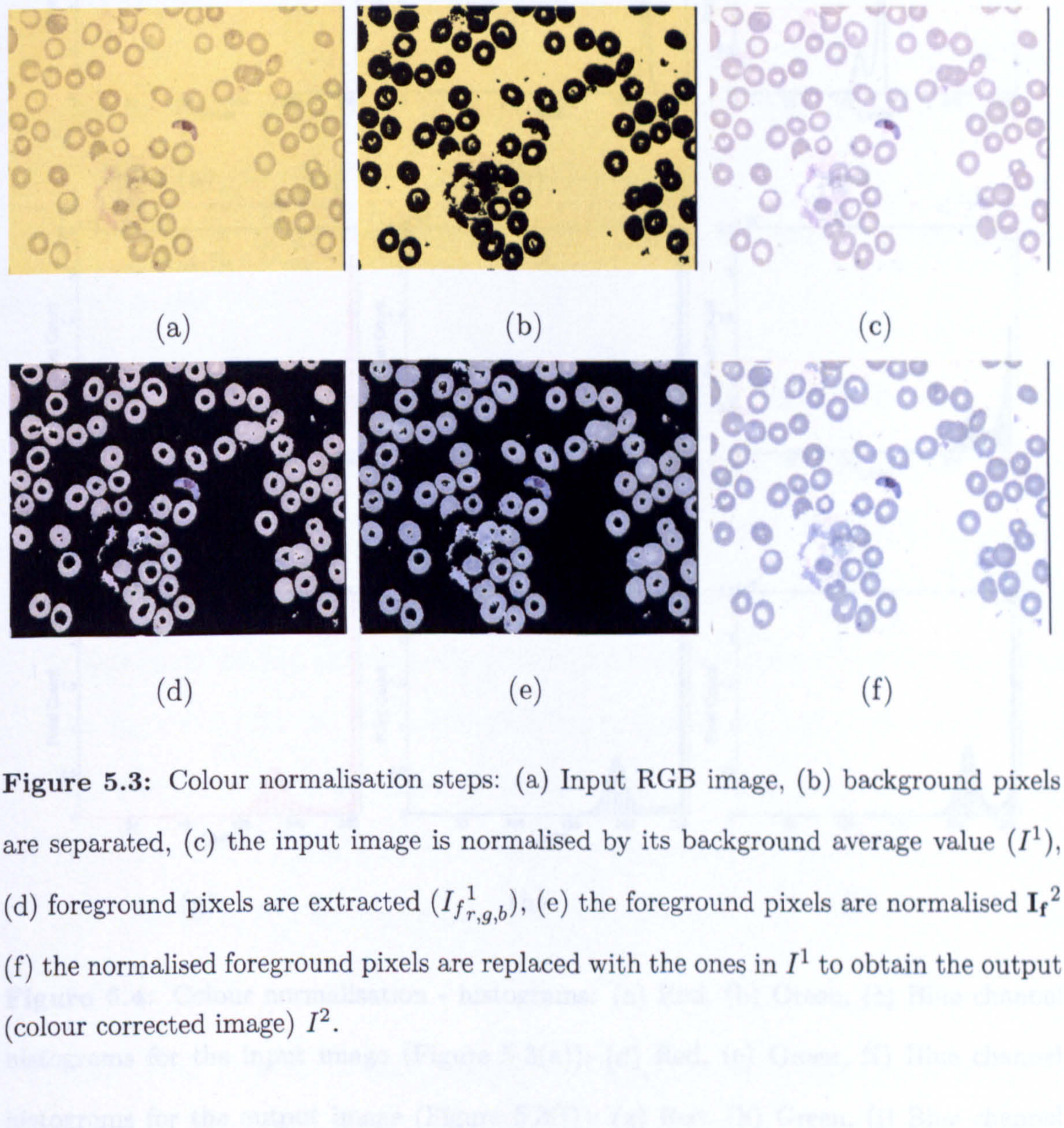
4. Transform only the foreground channels: $\mathbf{I_f}^2 = \mathbf{M^f I_f}^1$

5. Replace the foreground channels of $\mathbf{I^1}$ with $\mathbf{I_f}^2$ to obtain the final colour normalised output image $\mathbf{I^2}$.

The procedure is demonstrated with an example image in Figure 5.3. Following the image binarisation with technique explained in Section 4.3, the background pixels are extracted (Figure 5.3(b)). Using the background pixel averages and preset grey values ($G_i = \{255, 255, 255\}$ in (5.2), the input image is corrected (Figure 5.3(c)). Then from these image foreground pixels are extracted (Figure 5.3(d)) and corrected according to the grey values determined by the reference set (Figure 5.3(e)). The final result is obtained by replacing the normalised foreground pixels (Figure 5.3(f)). Figure 5.4 demonstrates the changes in the histograms of each channel during the process. It can be seen that as the result of the corrected process, the output histograms are shifted (Figure 5.4(d), 5.4(e),5.4(f)) and aligned to reference set histograms (Figure 5.4(g), 5.4(h),5.4(i)). It is also observable that the histograms in the output have zero values in some indexes which are introduced by the scaling and replacing of the pixel values to integer histogram bins. In fact, these entries (zero) are shifted and accumulated in the nearest bin. The same effect is observable in the reference set histograms since reference images are also normalised with respect to their background average values and preset grey values ($G_i = \{255, 255, 255\}$).

5.3 Experimental Results

As described by the procedure, the colour normalisation is done in two steps. In the first step, the input image channels are corrected using the ratio of their background average values to the preset background values $G_i = 255, 255, 255$. In the next step, the resulting image foreground channels are corrected using the ratio of their average values to the ones calculated from a reference set. A reference set (including 12 images) was prepared from the images supplied by NIMR. However, the images were not chosen in terms of quality. Only criterion is that the images have to be acquired with the same



(a) (b) (c)

(d) (e) (f)

Figure 5.3: Colour normalisation steps: (a) Input RGB image, (b) background pixels are separated, (c) the input image is normalised by its background average value (I^1), (d) foreground pixels are extracted ($I_{f^1_{r,g,b}}$), (e) the foreground pixels are normalised $\mathbf{I_f}^2$ (f) the normalised foreground pixels are replaced with the ones in I^1 to obtain the output (colour corrected image) I^2.

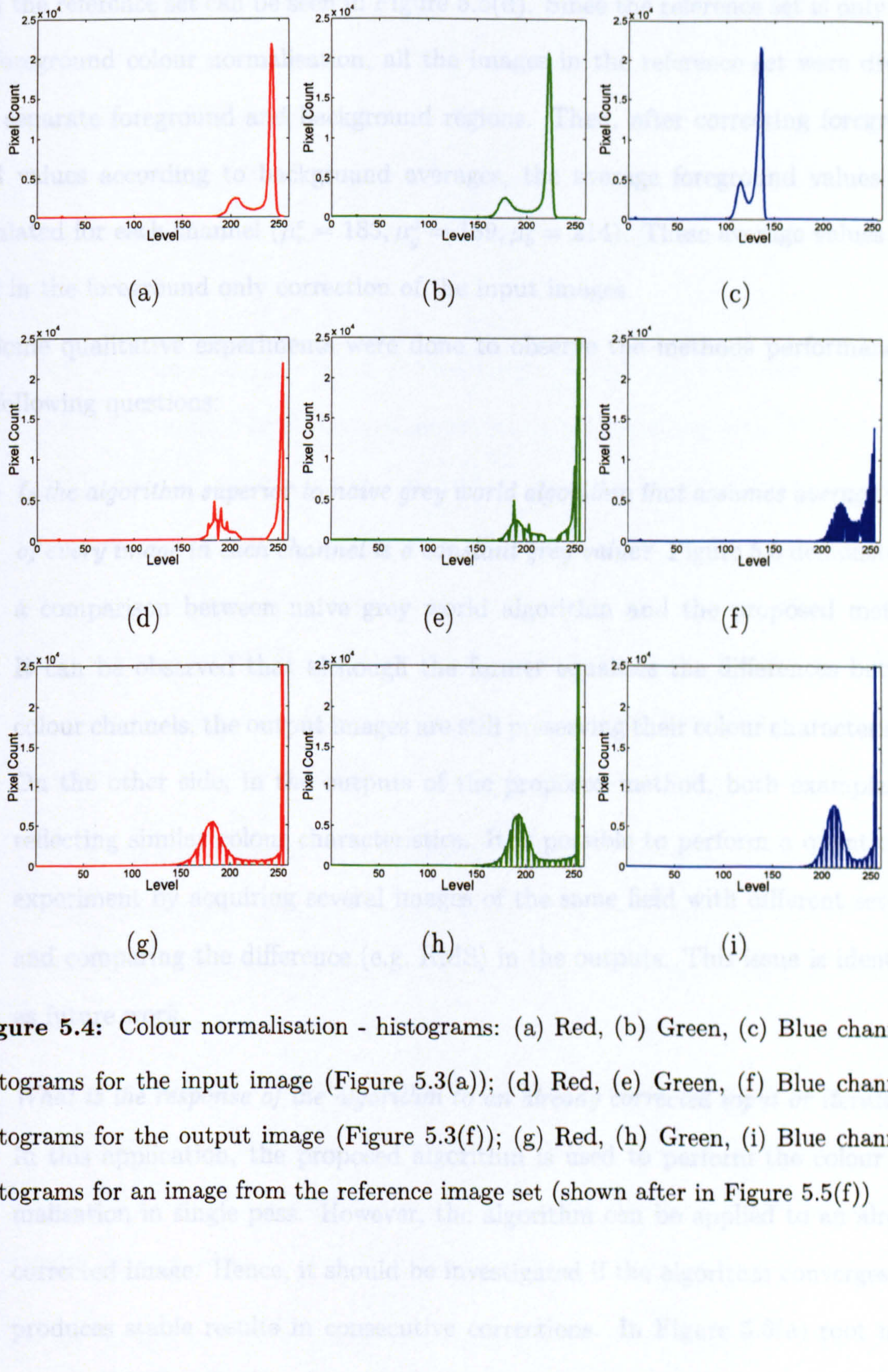


Figure 5.4: Colour normalisation - histograms: (a) Red, (b) Green, (c) Blue channel histograms for the input image (Figure 5.3(a)); (d) Red, (e) Green, (f) Blue channel histograms for the output image (Figure 5.3(f)); (g) Red, (h) Green, (i) Blue channel histograms for an image from the reference image set (shown after in Figure 5.5(f))

illumination and camera settings. The illuminant or camera settings are not known and required. This is because the aim here is to correct the images to reflect the same colour characteristics instead of obtaining the illuminant characteristics. An image from the reference set can be seen in Figure 5.5(d). Since the reference set is only used for foreground colour normalisation, all the images in the reference set were divided into separate foreground and background regions. Then, after correcting foreground pixel values according to background averages, the average foreground values were calculated for each channel ($\mu_r^c = 183, \mu_g^c = 189, \mu_b^c = 214$). These average values were used in the foreground only correction of the input images.

Some qualitative experiments were done to observe the methods performance to the following questions:

- *Is the algorithm superior to naive grey world algorithm that assumes average value of every image in each channel is a constant grey value?* Figure 5.5 demonstrates a comparison between naive grey world algorithm and the proposed method. It can be observed that although the former equalises the differences between colour channels, the output images are still preserving their colour characteristics. On the other side, in the outputs of the proposed method, both examples are reflecting similar colour characteristics. It is possible to perform a quantitative experiment by acquiring several images of the same field with different settings and comparing the difference (e.g. RMS) in the outputs. This issue is identified as future work.

- *What is the response of the algorithm to an already corrected input or iterations?* In this application, the proposed algorithm is used to perform the colour normalisation in single pass. However, the algorithm can be applied to an already corrected image. Hence, it should be investigated if the algorithm converges and produces stable results in consecutive corrections. In Figure 5.6(a) root mean square (RMS) difference is plotted versus the consecutive iterations. It can be

observed the significant difference is obtained in the first run, after few iterations algorithm converges as the ratios (i.e. diagonal entries of M) approach to 1. In Figure 5.6(b), 5.6(c) foreground and background scaling ratios are plotted versus the iteration number, respectively. It can be observed that the both of the ratios are synchronised with RMS; the foreground ratios converge to 1; however, the background ratios stabilise above 1. This is due to truncation of values to 255 that are exceeding 255. Thus, the background average value of the image can not reach average 255 unless all background pixels have the same value at the start of the iteration.

- *Is the algorithm robust against different input colour characteristics?* We have applied the method to all images in our image set including over 700 images acquired under different unknown conditions. Some of the images were taken from UoW microscope-camera setup, some of them were provided by NIMR as explained in Chapter 3. It was observed that the method is robust against different input characteristics. Some examples of input-output pairs are shown in Figure 5.7. Note that the input images have clearly different colour characteristics. The outputs show that the proposed colour normalisation is quite successful.

- *What is the effect of non-uniform illumination?* The colour normalisation is performed globally for the input image. Hence, the illumination is assumed to be uniform. Figure 5.8 demonstrates example input-output pairs when the illumination is not uniform. It can be observed that there are some errors in the outputs. However, the errors in these examples are due to foreground background segmentation rather than the colour normalisation process itself. If the illumination is not uniform one of the two methods described in Section 3.1 can be used as a pre-processing step.

- *Is the method specific to malaria or any special set of images?* The method has been developed for peripheral thin blood film images. It assumes the image

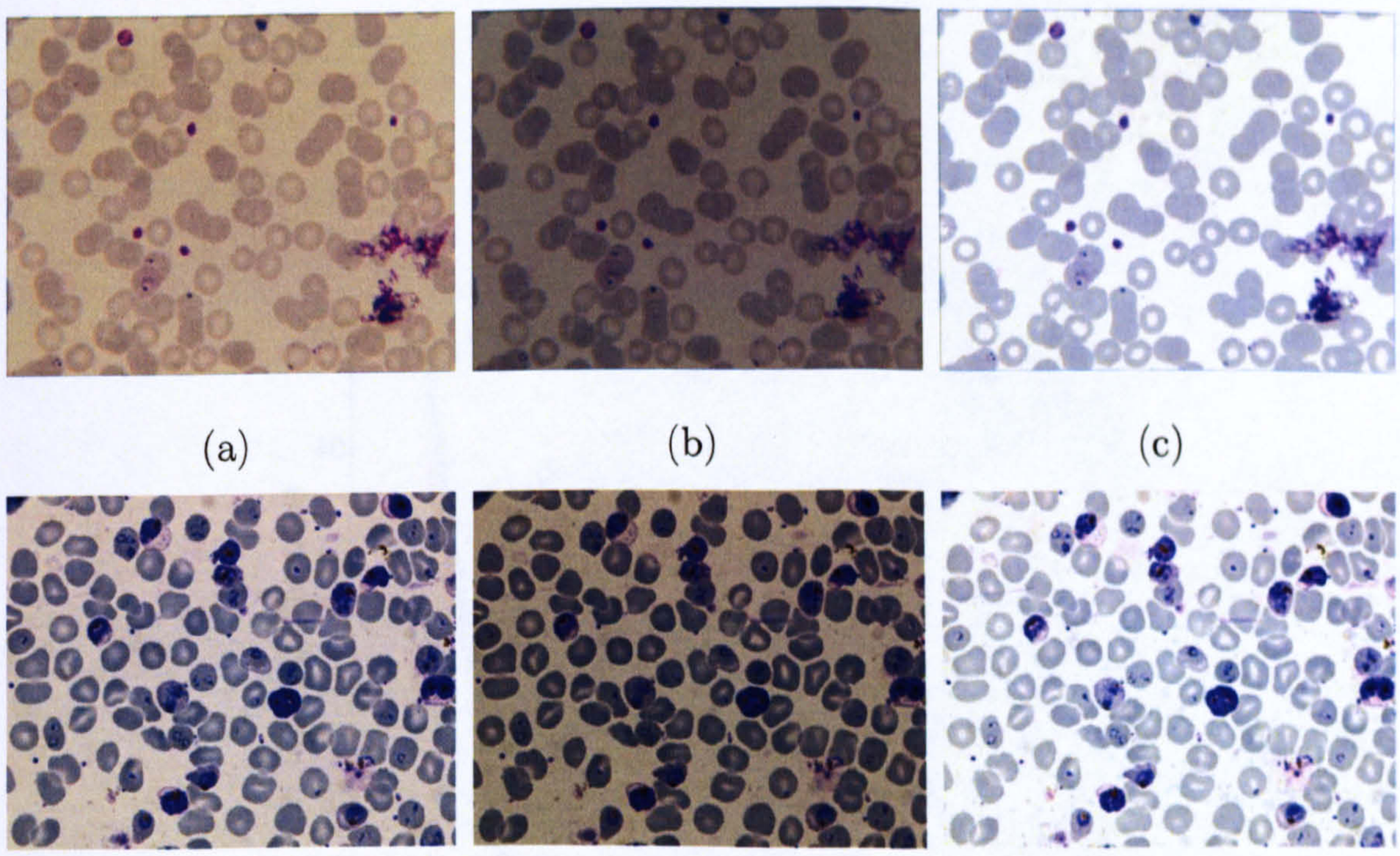
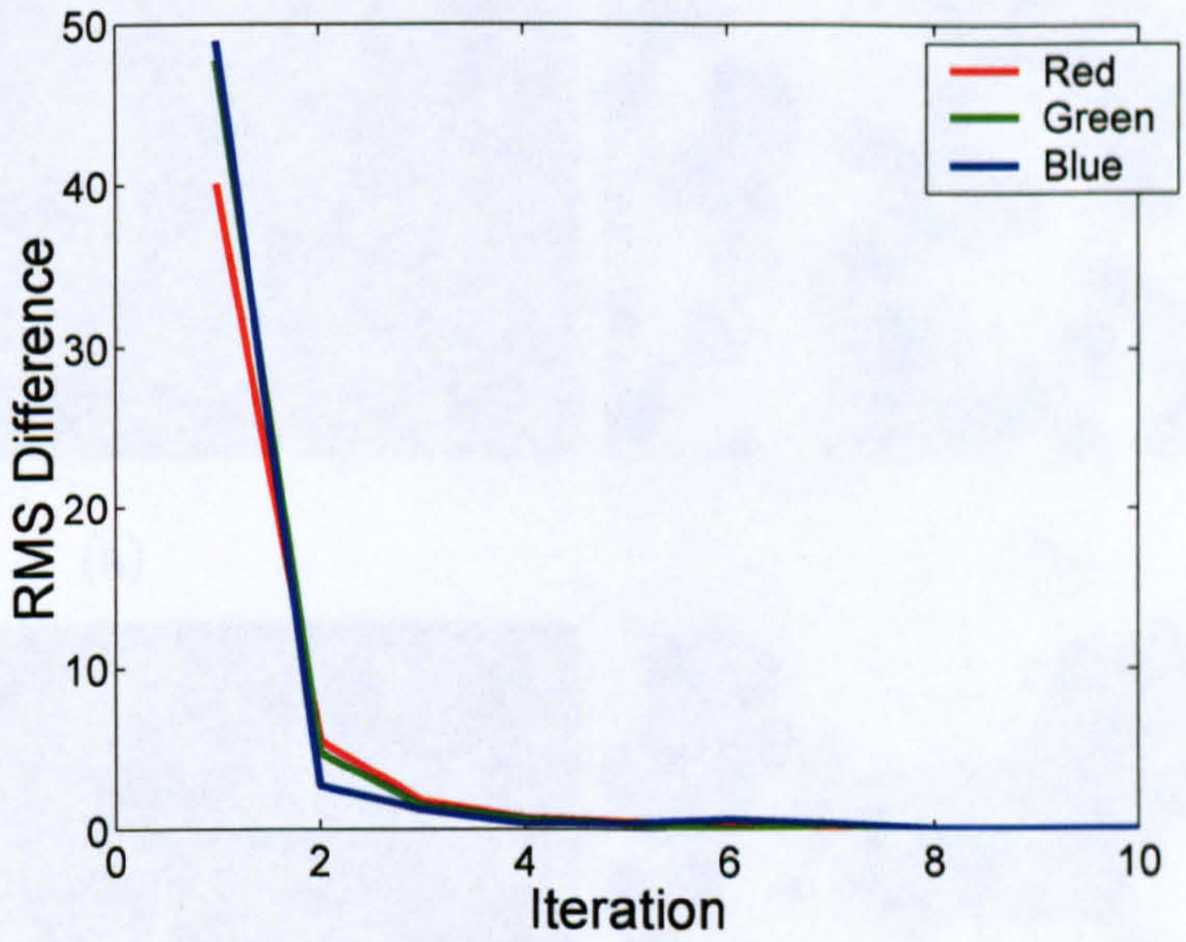


Figure 5.5: Colour normalisation: Comparison of the method with naive grey world algorithm ($G_i = \{128, 128, 128\}$) (a) An input image (b) normalised by naive grey world algorithm (c) normalised by the proposed method. (d) An input image from the reference image set, (e),(f) the results are in the same order.

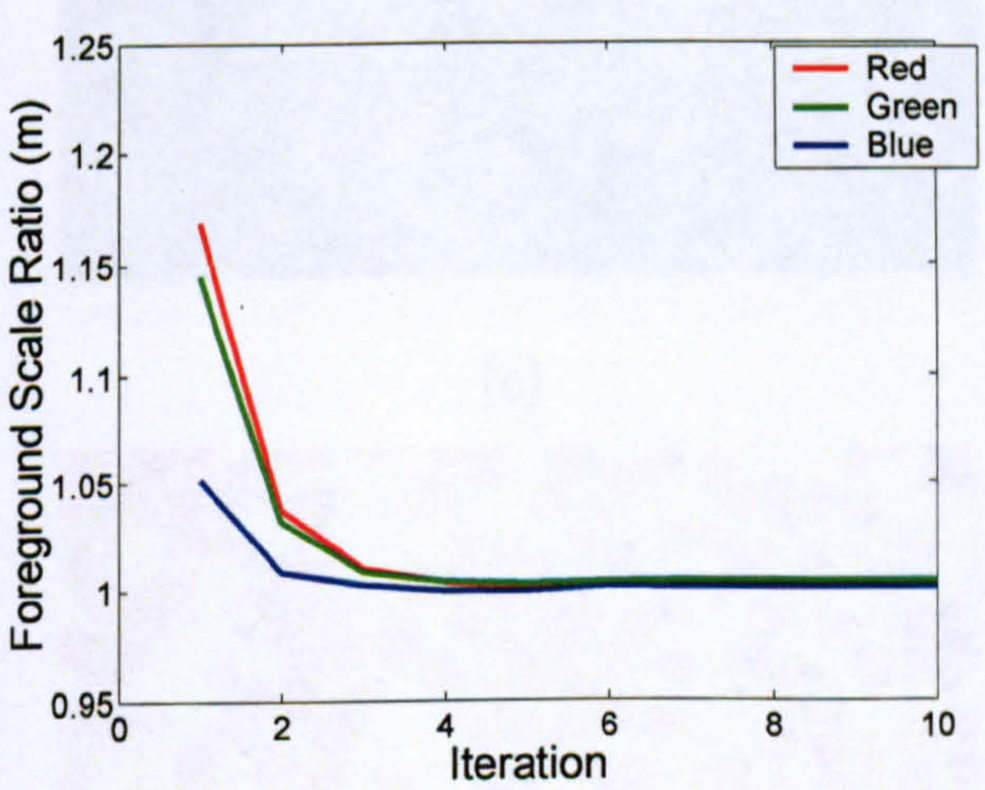
is separable into two foreground and background regions: that the background region must be plasma and the foreground region must contain mainly red blood cells (because of the reference characteristics). To demonstrate the robustness of the algorithm, some random images of other blood parasite types were collected from World Wide Web using the Google search engine. Figure 5.9 demonstrates some example input-output pairs. The outputs show that the method is very effective.

5.4 Conclusions

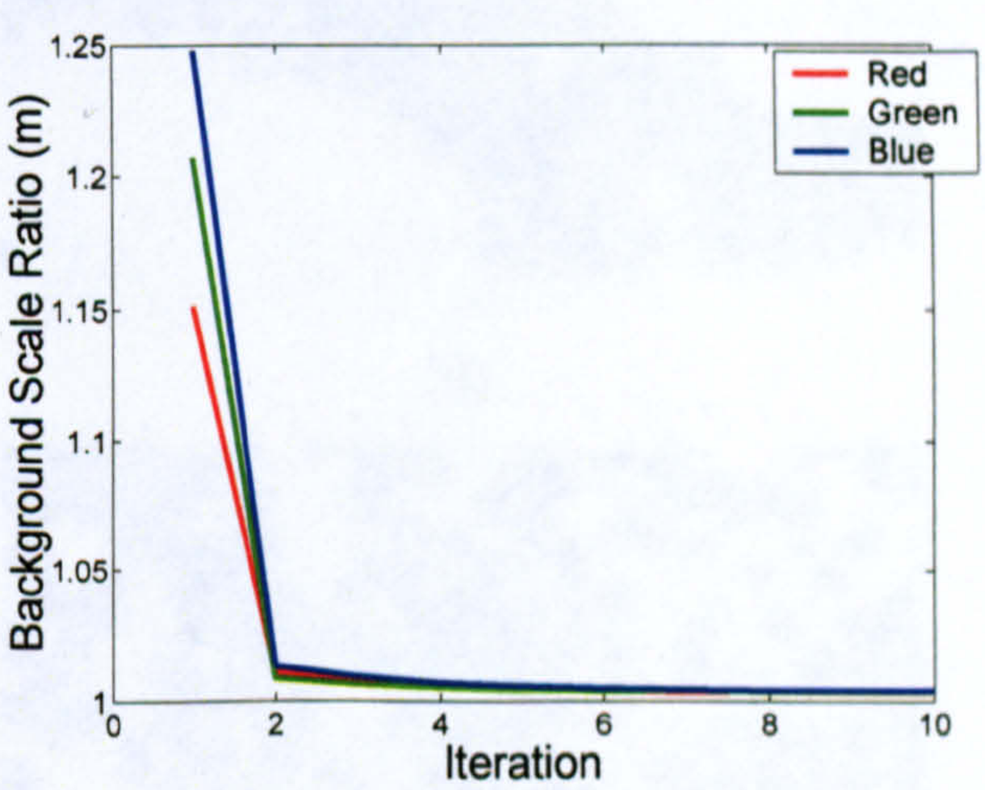
A new colour normalisation method has been presented for peripheral thin blood film images. The method exploits the special characteristics of the images which are separa-



(a)

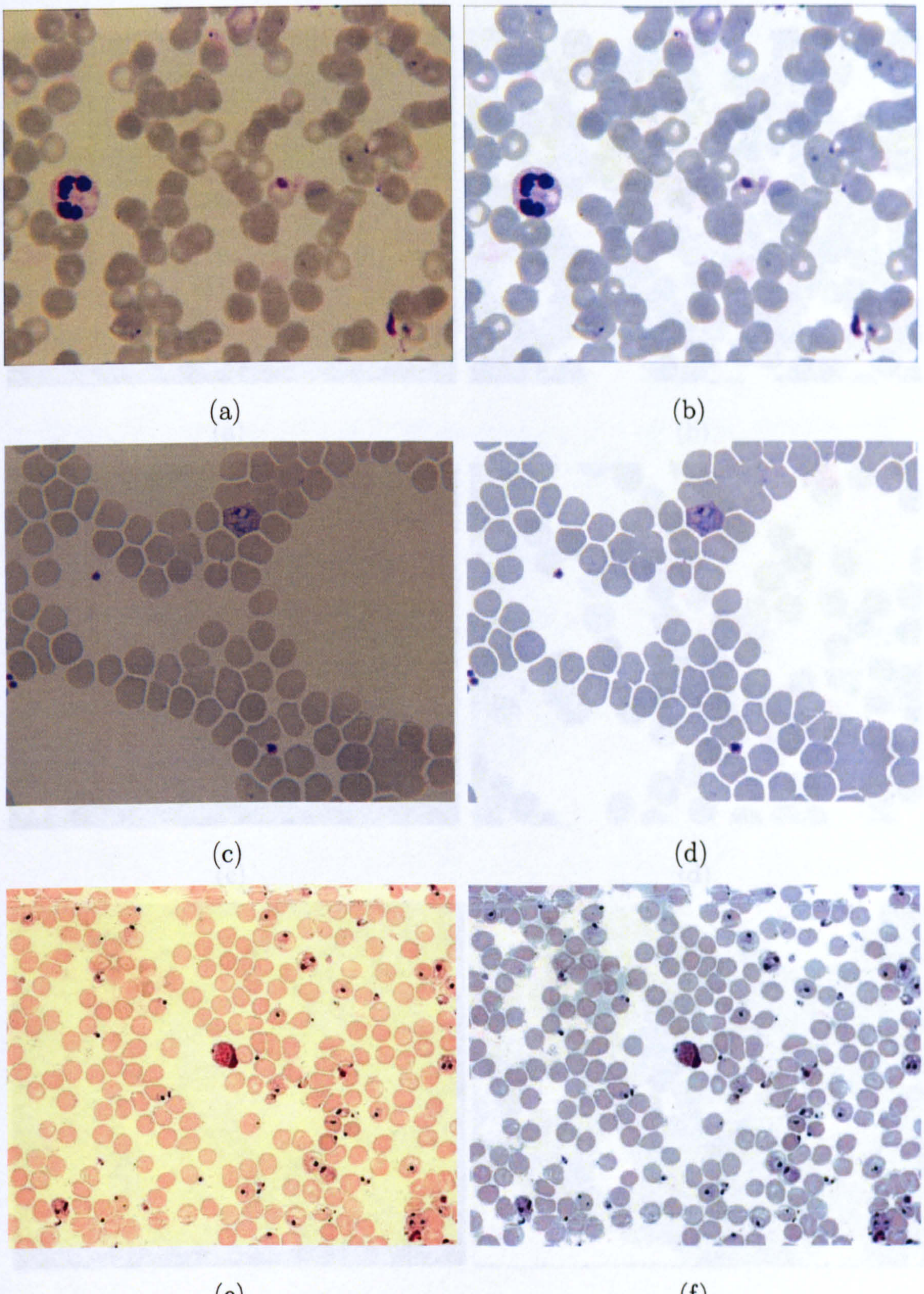


(b)



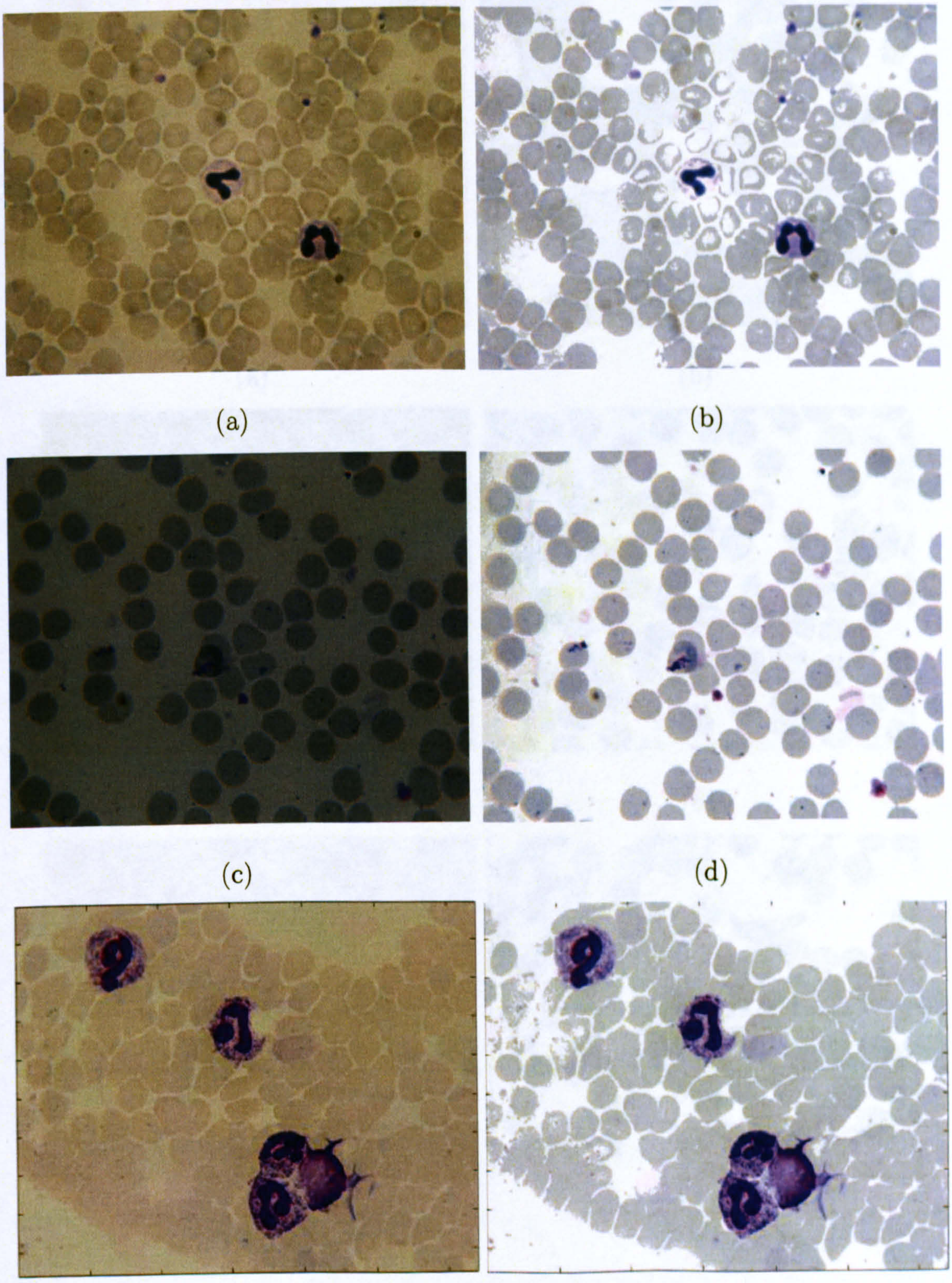
(c)

Figure 5.6: Convergence of the colour normalisation algorithm: (a) RMS difference between input and output, (b) foreground scaling ratio (m), (c) background scaling ratio versus iterations.



(a) (b) (c) (d) (e) (f)

Figure 5.7: Colour Normalisation Examples: Some example input-output pairs from ((a)-(d)) UoW and (e)-(f) NIMR image sets.



(a) (b) (c) (d) (e) (f)

Figure 5.8: Colour normalisation: effect of non-uniform illumination.

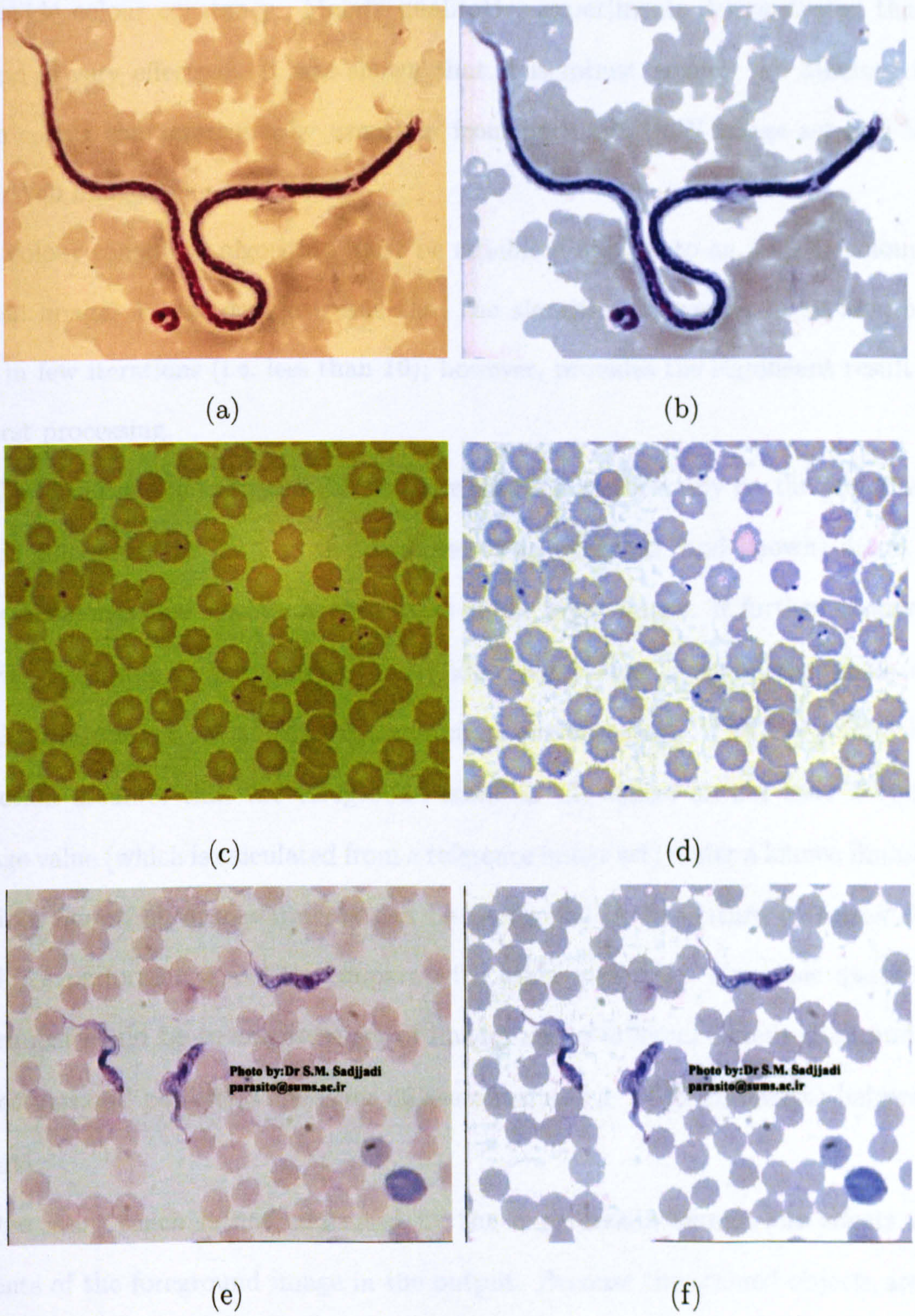


Figure 5.9: Colour normalisation - example images from other sources: Images from WWW/Google search engine, input-output pairs show (a)-(b): Wuchereria bancrofti (a Filariasis parasite) (c)-(d) P. Falciparum (e)-(f) Trypanosoma cruzi (i.e. Chagas Disease).

ble to foreground and background regions. Then it utilises the grey world assumption to provide colour constancy. Mainly qualitative experiments demonstrated that the method is very effective. It was shown that it is robust against the different input image colour characteristics by examples from both the UoW image set and World Wide Web images.

A colour constancy algorithm must be reliable if applied to an already colour normalised image. It was demonstrated that the algorithm converges to an idempotent state in few iterations (i.e. less than 10); however, provides the significant result after the first processing.

The basics of the proposed colour normalisation method rely on the fact that the plasma colour can be used as the response of an expected (and known) scene. The method assumes that plasma is observable in the input image. It further assumes the plasma (background) is colourless under ideal conditions; it scales the whole image to shift the plasma colour towards maximum possible scale. The foreground image correction assumes that the foreground scene in the image should have a constant average value (which is calculated from a reference image set) under a known illuminant.

More quantitative experiments can be performed in the future to assess the algorithm's performance and to compare with other methods. The basic quantitative experiment could be to acquire a set of images under different illumination and camera settings and perform some error measurements (e.g. RMS difference) between the outputs.

One issue which is not addressed by the experiments here is the effects of the contents of the foreground image in the output. Because the stained objects are (e.g. parasites, white blood cells, platelets, and artefacts) usually darker than red blood cells, an image having an excessive number of stained objects may shift the average value of the foreground to lower values. Therefore, the content of the reference foreground images that are used in calculation of the averages should be similar to the input image contents. For this reason, the proposed method can not be directly applied to thick

film images, where foreground objects are not guaranteed to be present and RBCs are not observable at all.

The necessity of the colour normalisation came from the colour based classifications of the stained objects that are introduced later. In the following chapter, the effect of the colour normalisation on the stained pixel classification process will be quantitatively evaluated and demonstrated. This method was presented by the author in [39].

Chapter 6

Stained Object Extraction

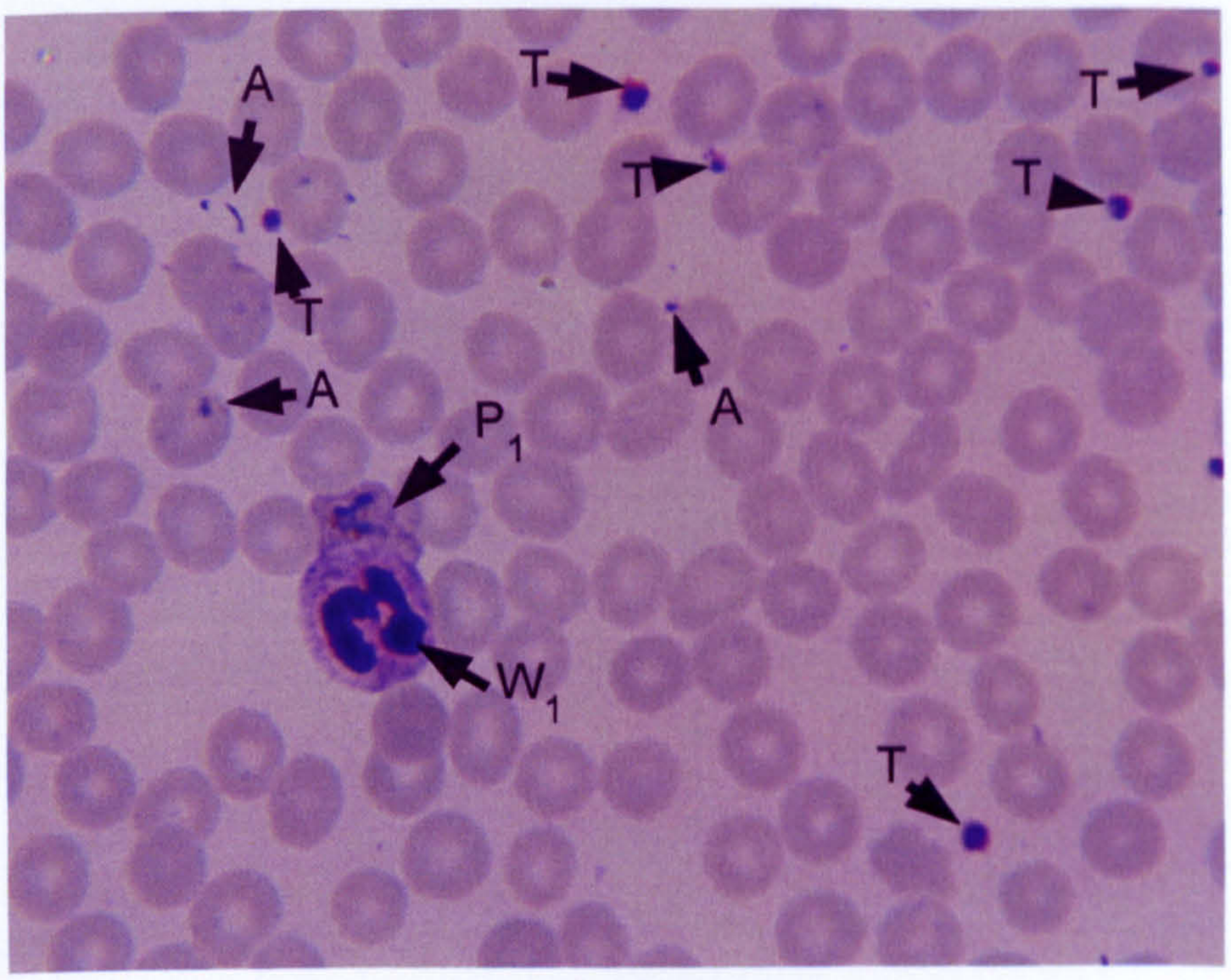
6.1 Introduction

Microscope examination of blood is usually performed on stained samples. Staining is a chemical process which highlights the object of interest in the sample; usually it is specific to the application. The common staining technique used for malaria diagnosis in peripheral blood thin blood films is the Giemsa-stain; however, in thick films Field's staining technique is preferred due to its speed. Both these techniques belong to a larger group that is referred to as "Romanovsky" stains which are a mixture of polychrome methylene blue and eosin [111]. In thin films the blood cells are fixed with methanol prior to the staining. Hence, the membranes of the blood cells are preserved. As explained in the previous chapter, depending on the imaging conditions, there can be various differences in the observed colours of the components. However, if examined using a tungsten light microscope (which is common practice), the components that have a nucleus (i.e. White Blood Cells (WBCs)), parasites, and some artefacts) will appear dark blue-purple, and platelets will have a saturated pink colour if they are stained. However, the colours vary according to the lighting and imaging conditions.

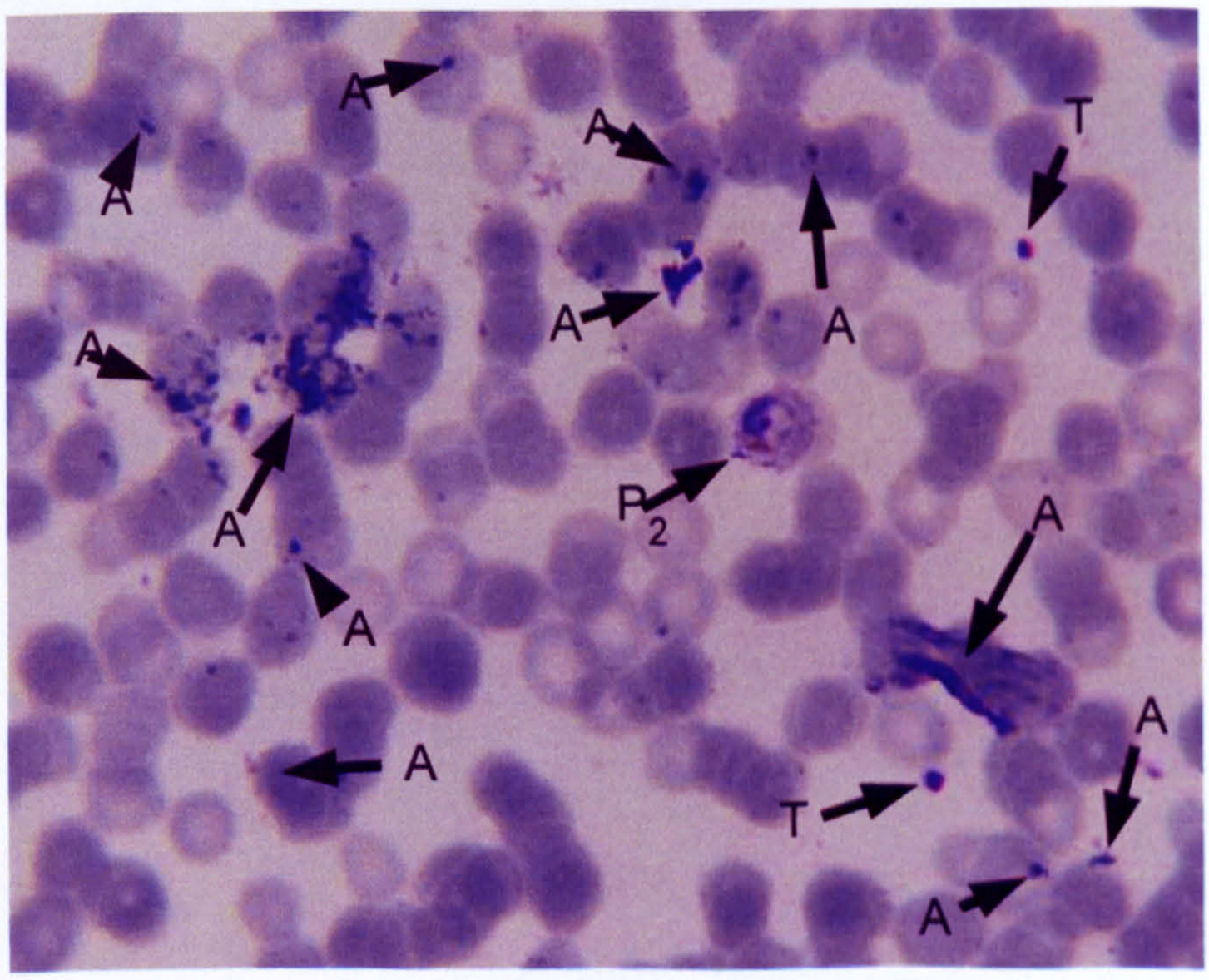
It is appropriate to clarify the terms *stained pixels* and *stained objects* before embarking on the technical discussions. A stained pixel is a pixel in the image which has distinctive, saturated colour compared to the colour of the Red Blood Cells (RBCs). A stained object can be roughly defined as a connected stained pixel group and the body that includes the group (if any). In the case of an infected RBC, the stained object is ideally the whole RBC region which includes the stained pixels; in the case of a WBC, it is ideally the WBC nucleus and surrounding cytoplasm; in the case of a platelet, it is the only connected stained pixel group (since they do not have cytoplasm); in the case of an artefact that is caused by a RBC anomaly (e.g. iron deficiency, Howell-Jolly

body), it is the whole RBC region that includes the stained pixels; and in the case of an artefact that is related to other factors such as stain crystals, dust or bacteria it is the connected stained pixel group itself. However, there is a lot of diversity in the images that makes it very difficult to process the images (i.e. to extract stained objects) based on these expectations. Figures 6.1-6.4 show some example images that contain stained pixels. At this stage of the discussion, the categorisation of the stained objects is not important; however to visualize the above definitions, the stained objects on the images are marked such that: parasites, WBCs, artefacts, and platelets are shown with the letters (P), (W), (A), and (T), respectively. Note that, for the parasites (P) the parasite shape and maturity is not always the same, hence the infected RBC appearance varies from one to another; for the WBCs (W) the cytoplasm is not always visible and the shape varies from convex W_4 to complex W_6; some of the platelets (T) are isolated on the plasma while some of them are touching the RBCs; and artefacts appear randomly and can be on top of the RBCs. However, the stained pixels have to be found and the stained objects have to be extracted and labelled. This process is largely unrelated to the blood component which they originate from.

In Chapter 4, it was explained that in Computerised Diagnosis of Malaria (CDM) from thin blood film images both deductive and inductive approaches can be followed. In [34], an inductive strategy was followed. First the stained pixels were detected; then the objects (i.e. RBCs) which are covering the detected stained pixels were found by morphological operations and passed to the parasite life-stage analysis. In [35], the approach was first to segment and locate the objects (i.e. RBCs) within the image; then after a global threshold to roughly locate the cells that contain stained pixels, the inner regions of the cells were analysed further to extract the stained regions. Both of the methods employed a thresholding operation to detect stained pixels. In [34] the stained pixels were detected by histogram thresholds on the top-hats of the Hue and Saturation channel images. The problem with this method is that even when the image does not contain any stained pixels the regional maxima and the histogram-based

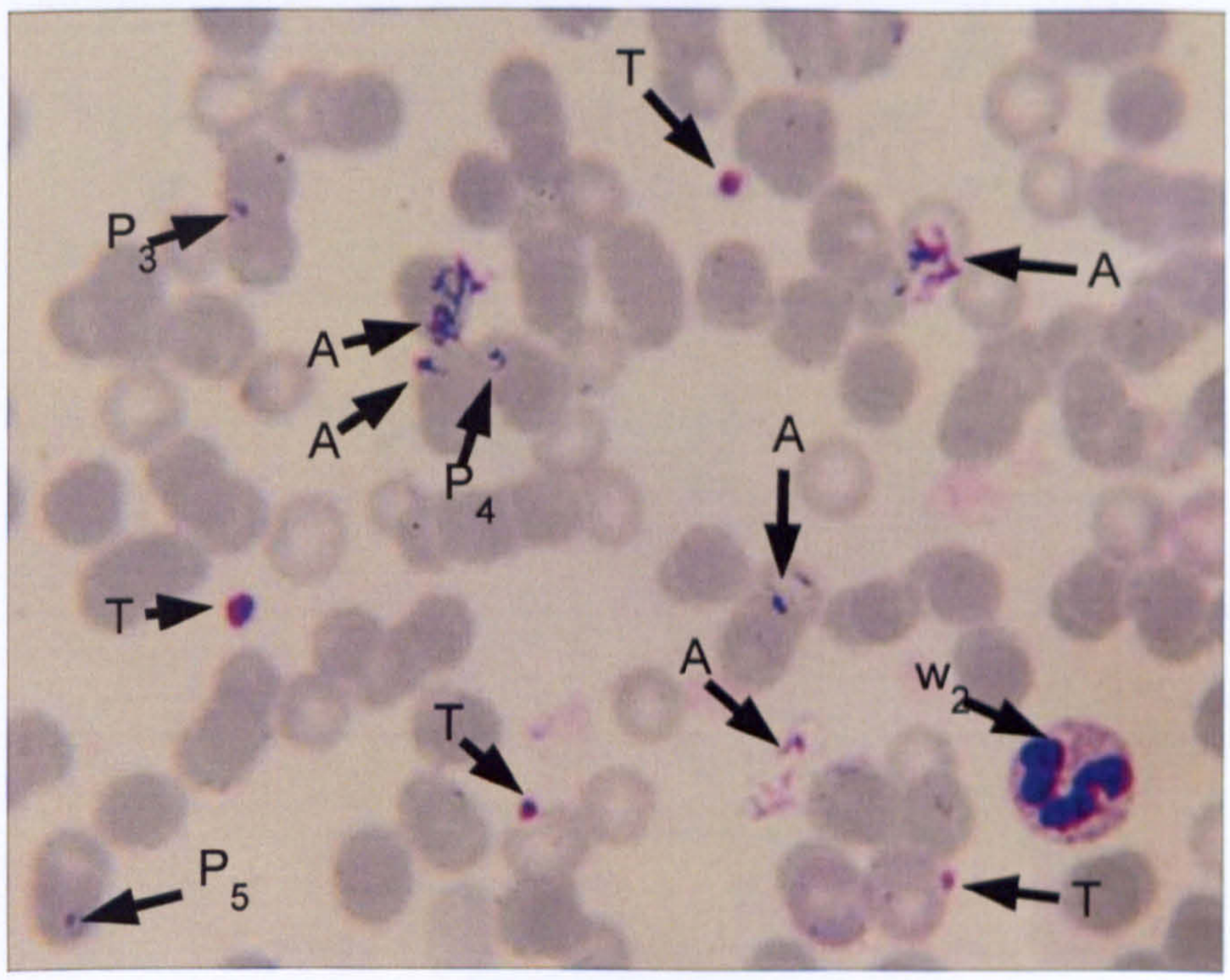


(a)

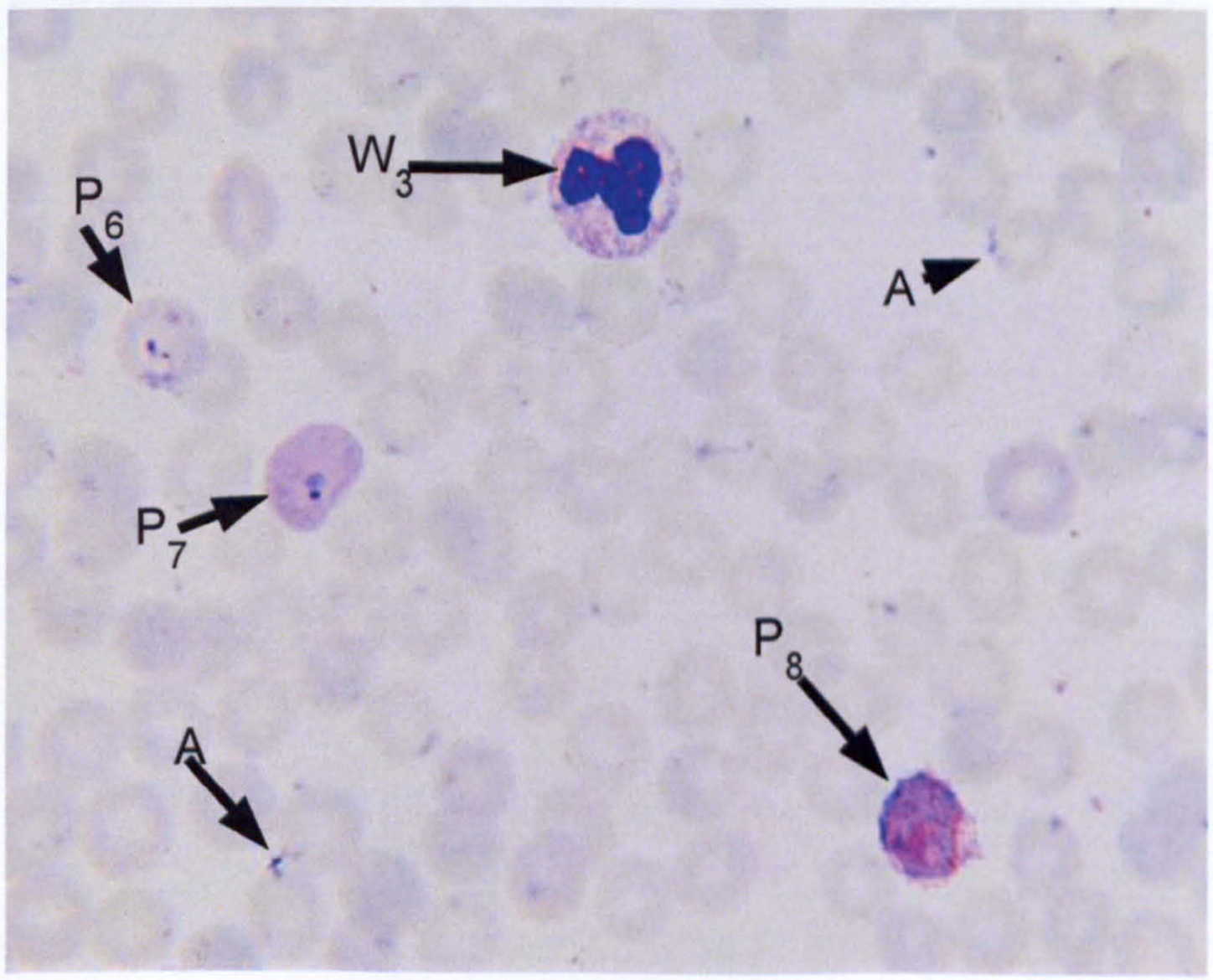


(b)

Figure 6.1: Stained object examples: (P) parasite, (W) white blood cell, (T) platelet, (A) artefacts

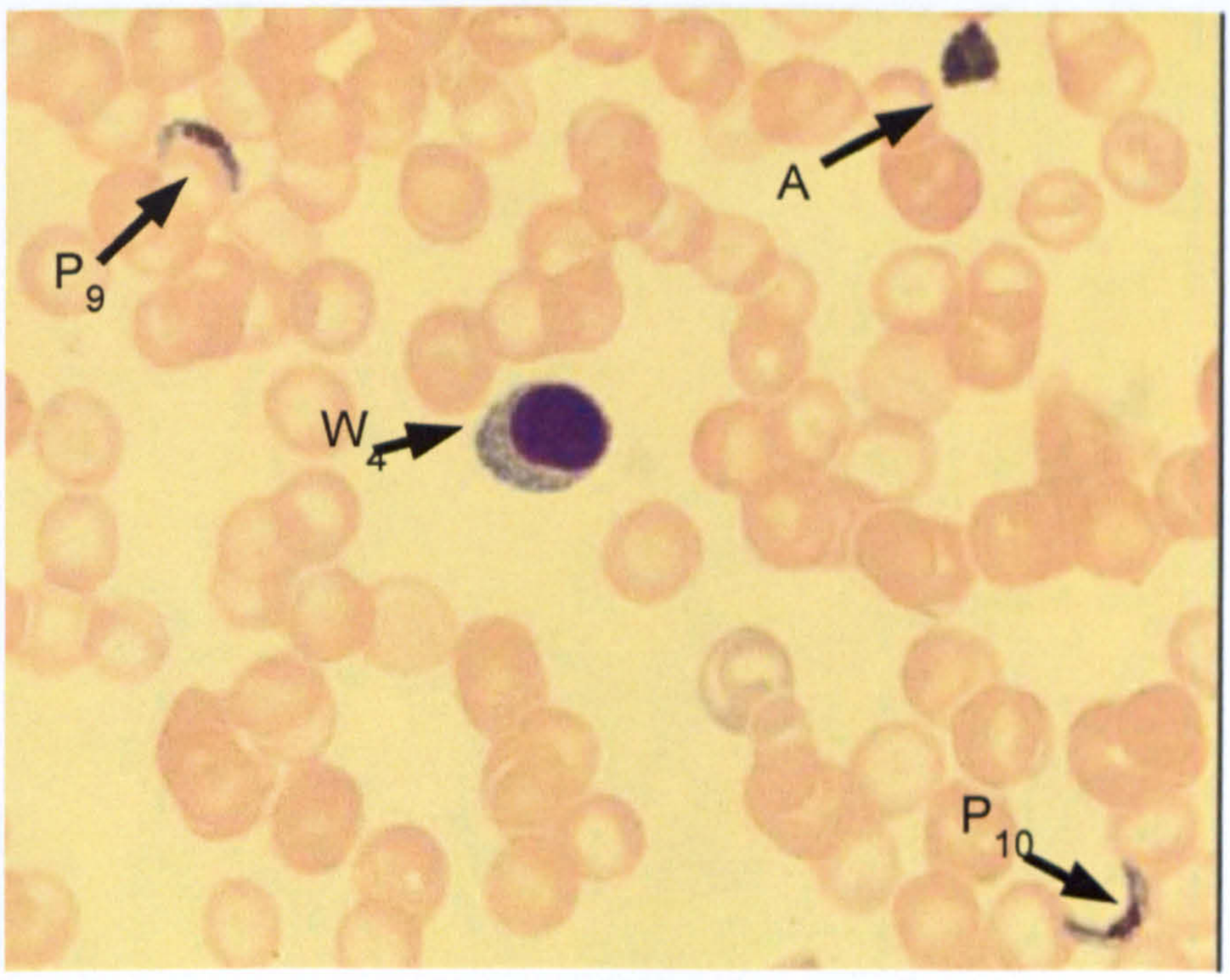


(a)

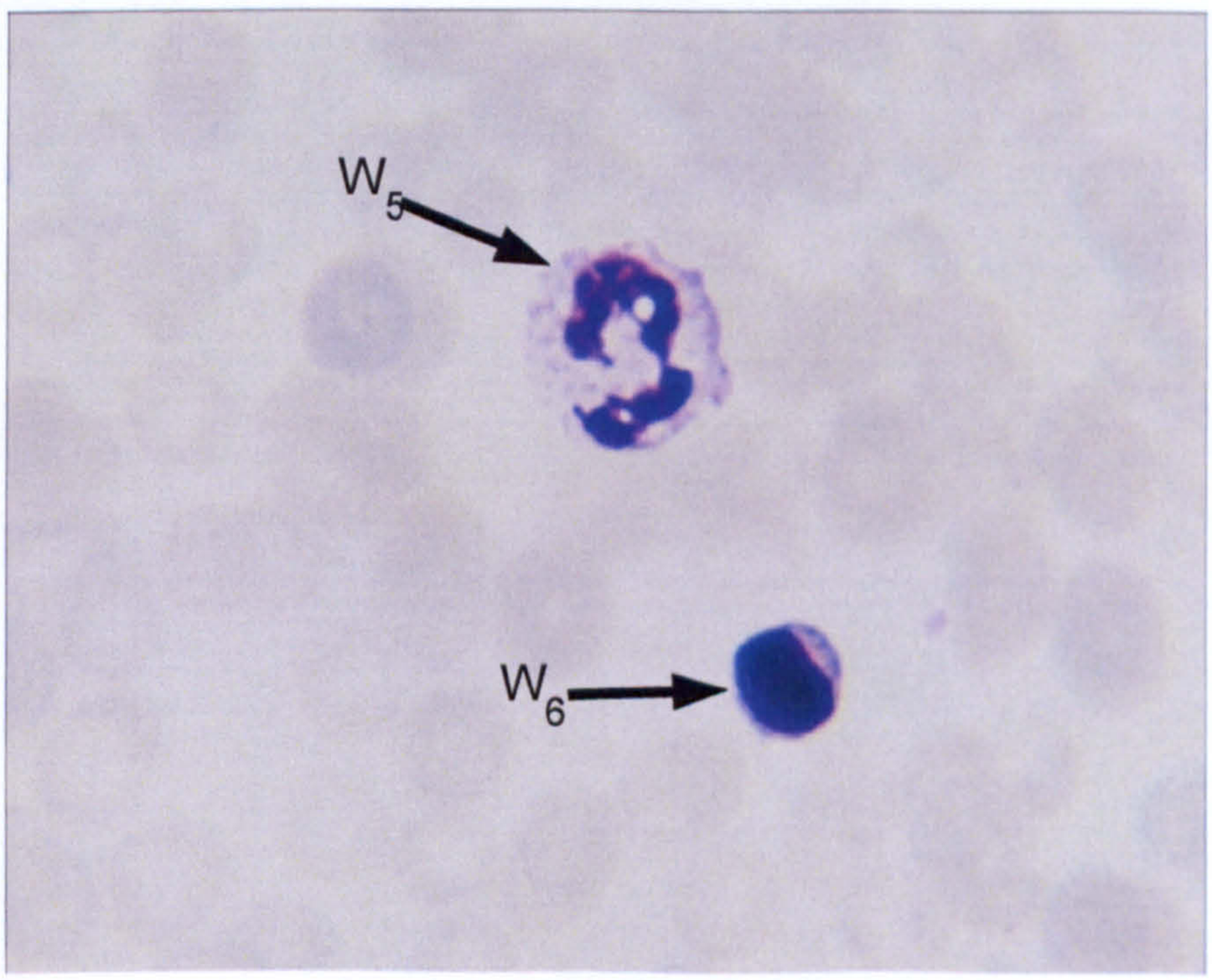


(b)

Figure 6.2: Stained object examples: (P) parasite, (W) white blood cell, (T) platelet, (A) artefacts

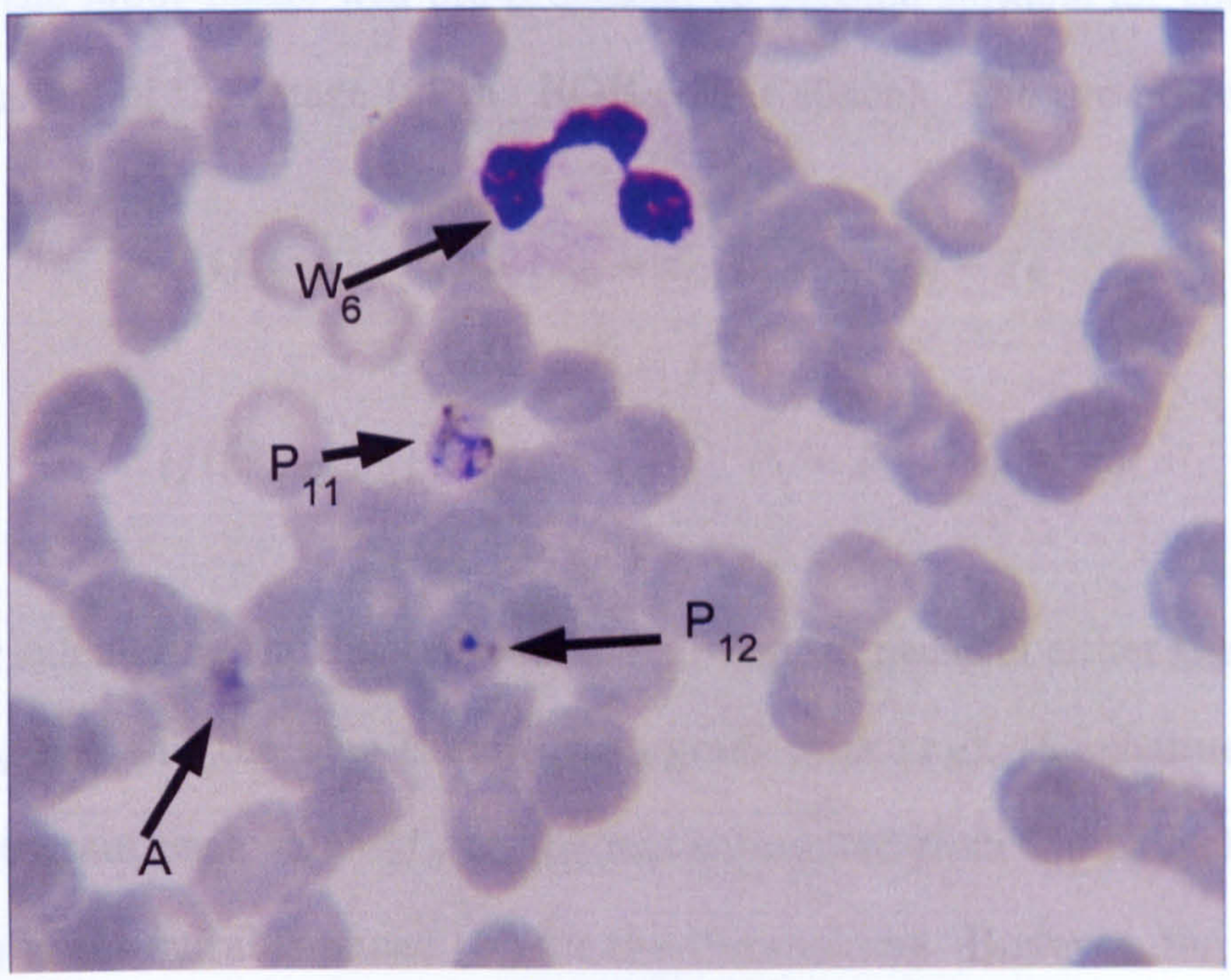


(a)

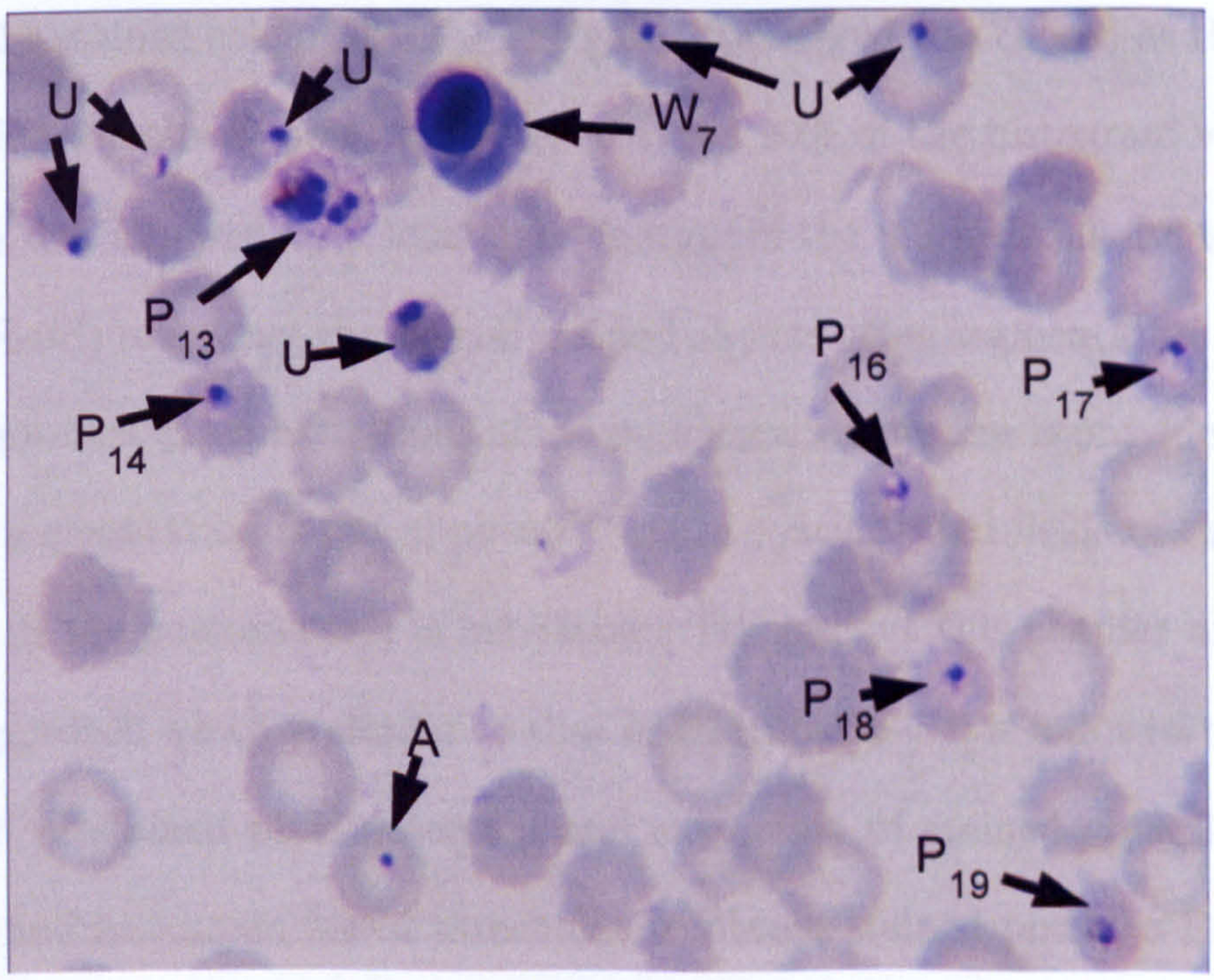


(b)

Figure 6.3: Stained object examples: (P) parasite, (W) white blood cell, (T) platelet, (A) artefacts



(a)



(b)

Figure 6.4: Stained object examples: (P) parasite, (W) white blood cell, (T) platelet, (A) artefacts, (U) unknown

threshold operations can extract some pixels as stained pixels. In [35] this problem was addressed: the input image was pre-processed to become suitable for thresholding in the blue channel of the image (in the RGB colour space). The pre-processing was as follows:

$$if((x_B - x_R) > (x_B - x_G)) \quad P = f_B - f_R$$
$$else \quad P = f_B - f_G \tag{6.1}$$
$$if((x_B - x_R) > (x_B - x_G)) \quad P_1 = P - (x_B - x_R)$$
$$else \quad P_1 = P - (x_B - x_G)$$

Thus by calculating the differences of the histogram peaks of either the red or green channel (x_R, x_G) and the blue channel histogram peak (x_B), the channel having the farthest histogram peak value (f_R or f_G) was subtracted from the blue channel image (f_B) to obtain the suitable image (P) for the thresholding. However, the thresholding operation was performed on P_1 which is obtained after another subtraction operation to remove the stained colour "bias". The global threshold was chosen as the histogram index that is the halfway point between 10% and 90% of the histogram volume of P_1. The process was followed by a local examination of the marked cells (obtained by the global threshold) to extract the refined stained objects using segmentation information.

A discussion on the constrictions of the segmentation process is provided in Chapter 4. Because segmentation is not applicable to the concentrated fields of the thin blood films, error in the post-analysis is inevitable. Throughout this chapter an alternative inductive approach which is similar to that of Di Ruberto [34] is followed to investigate the problem of stained pixel detection and extraction of stained objects. Instead of histograms and histogram-based thresholds to obtain stained pixels, a Bayesian pixel classifier is employed. In order to eliminate the necessity of segmentation information (as discussed in Section 4.8), the stained objects were extracted with local morphological operations. The colour normalisation method introduced in the previous chapter is used here to improve the efficiency of the Bayesian pixel classifier. A detailed evaluation was performed using the stained pixel ground-truth data obtained by manual

thresholding as described in Section 3.2.2.

The main contributions introduced in this Chapter are:

- A generalised solution is proposed for the problem of stained pixel detection by using a supervised learning scheme instead of relying on assumptions such as "stained pixels are darker than non-stained pixels". Additionally, there are no assumptions about the morphology of the objects in the image. The application to thick films is possible. Moreover, it is possible to apply our solution to other staining methods and microscope examination tasks by simply replacing the RGB samples that are used in the training.

- The stained object extraction method following the stained pixel detection task is specific to thin peripheral blood film images. It eliminates the necessity of pre-segmentation of the thin film images by proposing an inductive strategy to extract stained objects. Hence, it prevents the stained objects from being divided.

- The methodology that is introduced in this Chapter is published by the author in [36, 37] with narrower contexts.

In the following two sections the Bayesian pixel classification and the stained object extraction processes are described. In Section 6.5 the evaluation experiments of the pixel classification performance and some results of the overall procedure are demonstrated. In the conclusion (Section 6.6), the advantages, disadvantages, and possible improvements are discussed.

6.2 Stained Pixel Detection - Bayesian Classifier

In contrast to the histogram-based thresholding operations which can be regarded as unsupervised classification schemes (i.e. clustering), the detection of the stained pixels can also be seen as a two class (stained, non-stained) supervised classification problem. Supervised classification requires a number of labelled (supervised) samples

and a training process to calculate a decision rule. In a problem where there are no observations (i.e. samples, features) the only decision rule could be based on *a priori* probabilities. Let, in a two class classification problem, w denote the state of nature such that $w \neq w_1 \rightarrow w = w_2$, and $w \neq w_2 \rightarrow w = w_1$ then the decision based on the *a priori* will be:

$$\text{Decide } w_1 \text{ if } P(w_1) > P(w_2); \text{ otherwise decide } w_2 \tag{6.2}$$

Bayesian decision theory, which is the fundamental theory for pattern recognition, suggests that if any features (i.e. observations, measurements) are present, the decision can be based on the *a posteriori* probabilities that minimise the decision error. For a two class (w_1, w_2) classification problem the Bayes decision rule for minimum error is expressed as follows [33]:

$$\text{Decide } w_1 \text{ if } P(w_1|\boldsymbol{x}) > P(w_2|\boldsymbol{x}); \text{ otherwise decide } w_2 \tag{6.3}$$

The $P(w_i|\boldsymbol{x})$ $(i \in \{w_1, w_2\})$ denotes the *a posteriori* probabilities: the probability of class w_i given the observation $\boldsymbol{x}$. Normally, these *a posteriori* are unknown, however according to Bayes' theorem they can be calculated as:

$$P(w_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|w_i)P(w_i)}{p(\boldsymbol{x})} \tag{6.4}$$

where $p(\boldsymbol{x}|w_i)$ is the class-conditional probability density function: probability of observing $\boldsymbol{x}$ given the class w_i. Because the $p(\boldsymbol{x})$, the probability of observing $\boldsymbol{x}$, is independent of the classes, the decision rule (6.3) can be rewritten as:

$$\text{Decide } w_1 \text{ if } \frac{p(\boldsymbol{x}|w_1)}{p(\boldsymbol{x}|w_2)} > \frac{P(w_2)}{P(w_1)}; \text{ otherwise decide } w_2 \tag{6.5}$$

Since the ratio of a priori probabilities is a constant value, in this form the decision rule is more focused on the left part of (6.5): the ratio of conditional probability density functions $p(\boldsymbol{x}|w_1)/p(\boldsymbol{x}|w_2)$ which is called the *likelihood ratio*. The Bayesian decision rule as in (6.3), (6.5) results in the minimum (average) classification error, in this sense

it is the optimal decision. The details can be read in [31–33]. However, there are some problems where the error in one class could be more important than the other. Hence, there is a slightly different form of (6.5) which considers the class-conditional risks (costs) to be taken into account to ensure minimum risk instead of minimum error.

$$\text{Decide } w_1 \text{ if } \frac{p(\boldsymbol{x}|w_1)}{p(\boldsymbol{x}|w_2)} > \frac{R_{12}}{R_{21}}\frac{P(w_2)}{P(w_1)}\text{; otherwise decide } w_2 \tag{6.6}$$

where R_{ij} is the cost of mis-classifying $\boldsymbol{x}$ as w_1 when $\boldsymbol{x} \in w_2$.

In stained/non-stained colour pixel classification, like most of the pattern classification problems, the *a priori* probabilities and the cost of mis-classifications are not known. Thus, the Bayesian classifier with a feature in the form of $\mathbf{rgb} = \{r, g, b\}$ colour vector can be formulated as follows:

$$\text{Decide } w_s \text{ if } L(\mathbf{rgb}) = \frac{p(\mathbf{rgb}|w_s)}{p(\mathbf{rgb}|w_{ns})} > \theta\text{; otherwise decide } w_{ns} \tag{6.7}$$

where w_s and w_{ns} denote the stained and non-stained classes, respectively; $p(\mathbf{rgb}|w_i)$ $(i \in \{s, ns\})$ denotes the class-conditional probability density function or more appropriately class-conditional probability mass function since **rgb** is formed of discrete (integer valued) variables [33]. θ is a threshold for the likelihood ratio: it represents implicitly the application dependent costs for the decision and especially the *a priori* probabilities of classes $P(w_s)$, $P(w_{ns})$ when they are not easily determinable.

6.2.1 Density Estimation Using Histograms

There are parametric and non-parametric approaches for estimating $p(\mathbf{rgb}|w_i)$. The parametric approaches tend to calculate parameters such as mean and covariance to represent the densities that are assumed to be in a specific distribution form (e.g. normal) [31,33]. However, non-parametric methods can be used to estimate densities without assuming a particular distribution form. In this study, a non-parametric method based on histograms is used [130]. The estimation of $p(\mathbf{rgb}|w_i)$ using histograms can be performed as follows:

1. Initialise (to zero) 3-d histograms for both stained $H_s(r,g,b)$ and non-stained $H_{ns}(r,g,b)$ classes. Count every occurrence of (r,g,b) to construct the histograms:

$$\begin{aligned} H_s(r,g,b) &= H_s(r,g,b)+1; \quad if \quad \mathbf{rgb}\in w_s \\ H_{ns}(r,g,b) &= H_{ns}(r,g,b)+1; \quad if \quad \mathbf{rgb}\in w_{ns} \end{aligned} \tag{6.8}$$

2. Normalise by the count of total samples (N_s, N_{ns}), respectively.

$$p(\mathbf{rgb}|w_s) = \frac{H_s(r,g,b)}{N_s} \quad p(\mathbf{rgb}|w_{ns}) = \frac{H_{ns}(r,g,b)}{N_{ns}} \tag{6.9}$$

A known problem with histogram-based density estimation is due to the limited number of samples. The problem space is scaled exponentially by the number of feature vector elements (i.e. curse of dimensionality [131]). In RGB colour space each colour channel is usually quantised to 256 discrete levels; that $p(\mathbf{rgb}|w_i)$ is required to have 256^3 distinct entries (i.e. bins). Hence, the number of training samples and the size of histograms are important parameters in the efficiency of the classifier.

Bayesian colour pixel classifiers that employ histogram-based estimation to obtain density functions are widely used in skin colour detection research, e.g. [63, 130, 132]. Usually, the likelihood ratio threshold θ value is obtained by Receiver Operating Characteristics (ROC) analysis. ROC analysis is used to compare different classification algorithms [48], efficiency of colour spaces [63], and also to determine the efficiency of the density estimation with respect to number of histogram bins [130].

6.3 Receiver Operating Characteristics

The ROC curve of a classifier is obtained by plotting its true detection (i.e. true positive) probabilities versus false detection probabilities (i.e. false positive) with respect to a control variable. ROC curve represents the trade-off between the true and false detection rates of a classifier; curves for different classifiers (or settings) can be compared to determine the optimal [48, 49].

For the proposed classification scheme, the control variable is the θ, the likelihood ratio threshold of the classifier; the true detection and false detection probabilities can be represented as $P(L(\mathbf{rgb}) > \theta|\mathbf{rgb} \in w_s)$, $P(L(\mathbf{rgb}) > \theta|\mathbf{rgb} \in w_{ns})$, respectively. The probabilities can be obtained by the ratio of true/false detection counts over total number of positive/negative samples.

In a ROC curve, the points $(0, 0)$ and $(100, 100)$ correspond to the classifier which makes all decisions as negative and positive, respectively. The points $(0, 100)$ and $(100, 0)$ correspond to the most and least desirable classifiers, respectively. The ROC plot of a random classifier follows the $(0, 0) - (100, 100)$ line. By calculating the areas under the ROC curves, a non-parametric comparison of different classifiers is possible: the curve that has the maximum area is the optimal provided that the curves are not crossing [31, 48]. If the curves are crossing, the optimal classifier can be searched in a limited interval according to the distribution of the samples between positive and negative classes.

In the experiments, we will use ROC analysis to determine the optimal number of histogram bins and likelihood ratio threshold θ of the classifier.

6.4 Stained Object Extraction

Stained pixel detection is a global process for an arbitrary image, and for all possible images since the likelihood ratio threshold (θ) is constant. The pixel classifier defines a stained pixel according to its colour pixel vector which is independent on the local characteristics of the image that the pixel is included. However, the stained objects are locally, contextually defined in that they may include non-stained pixels beside the stained pixels. It is difficult to develop any global assumption for the circularity, size or resolvable boundaries due to the artefacts (See examples in Figures 6.1- 6.4). Hence, the problem of stained object extraction is very challenging. It is also challenging because the assumptions based on the regular blood components (i.e. RBC, WBC

and platelet) fail due to the almost random appearance of the artefacts. From the CDM point of view stained object extraction has a role to define the object that is to be classified in the parasite/non-parasite classifier. Therefore, any enhancement may improve the classification performance. On the contrary, the errors introduced here, such as dividing a parasite into two parts, may complicate the classification (so the diagnosis).

The segmentation information was used in [35] to extract stained objects which were marked by the stained pixels. However, as was explained in Section 4.8 the segmentation information is not applicable to all the fields in a stained thin blood film. Here, a method is proposed which is an alternative to the use of segmentation information to extract stained objects.

6.4.1 Method

The stained/non-stained pixel probability scheme that was created for pixel detection is further utilised here. In the previous section, the likelihood ratio of a pixel $L(\mathbf{rgb})$ was calculated using the ratio of the w_s, w_{ns} density estimates (6.7). Let L_I represents the likelihood image for which all the corresponding pixel likelihood values in the image plane $I_{x,y}$ is calculated:

$$L_I(x,y) = L(I(x,y)) \quad \forall \{x,y\} \in \{I_{x,y}\} \tag{6.10}$$

where $I(x,y)$ is $\mathbf{rgb}$ colour vector and $L_I(x,y)$ is a real valued image in the range $[0,\infty]$. It would be practical to quantise $L_I(x,y)$ to be able to use it in local morphological operations. Let us assume that a function that maps the $L_I(x,y)$ value to a discrete valued and bounded range such as $[0,255]$:

$$L_I^q(x,y) = m(L_I(x,y)) \quad 0 \leq m(x) \leq 255 \tag{6.11}$$

The function $m(x)$ maps $([0,\infty))$ range to $(q = [0,\ldots,255])$. To do this, we first applied a logarithmic mapping, then assigned the result to the nearest integer in a

uniformly distributed interval (i.e. $[10^{-5}, 10^{+5}]$). Figure 6.5(a) demonstrates the likelihood image of the image shown in Figure 6.2, plot of the quantisation transfer function ($m(x)$) (Figure 6.5(b)), and the quantised likelihood image ($L_I^q(x,y)$) (Figure 6.5(c)).

Thus, we have obtained a discrete valued image of "*stained-ness*": a topological surface representation which can be used in morphological operations to extract stained objects locally. Note that, more than one stained pixel group may belong to the same stained object. Thus, the number of stained pixel groups may be smaller than the number of stained objects; the set of the stained pixels is contained in the set of the stained objects. Let $I_{x,y}$ define the image plane (as a set) and $S = \bigcup_{j=1}^{N} s_j$ the set of N stained pixels (or connected stained pixel groups) which are detected by the pixel classifier, the problem is to find $O = \bigcup_{i=1}^{M} o_i$ $(M \leq N)$ that S is a subset of $(S \subseteq O \subset I_{x,y})$. For example, if s is from a mature parasite as shown in P_8 in Figure 6.2(b), the stained pixel group may be equivalent to the object. When the parasite is immature it does not fill the cell completely, and depending on the species type it may be surrounded by some non-stained pixels (Figure 6.1(b) P_5, 6.2(b) P_6 and P_7).

Algorithm 3 (on page 133) provides a procedure to extract stained object using the stained pixels as markers.

The definitions of morphological operations are presented in Section 1.8. The white area top-hats (W_1 and W_2) extract the domes smaller than the chosen areas (α and β). However, they provide the binary sets that mark the domes when thresholded by 0. The reconstruction operator simply calculates the binary set that is connected to a particular pixel group (i.e. markers) among the others.

The algorithm suggests a heuristic solution for the problem by using average cell area as a basis. The stained pixel groups that are bigger than the average cell area value (estimated by area granulometry of the input image, see Section 4.2) are kept; the ones smaller than average cell area are further processed (Line 2). Line 3 gives the answer to the question "which is the set o_i that marks the dome of the likelihood image, that is limited by the average cell area and connected to stained pixel set s_j.

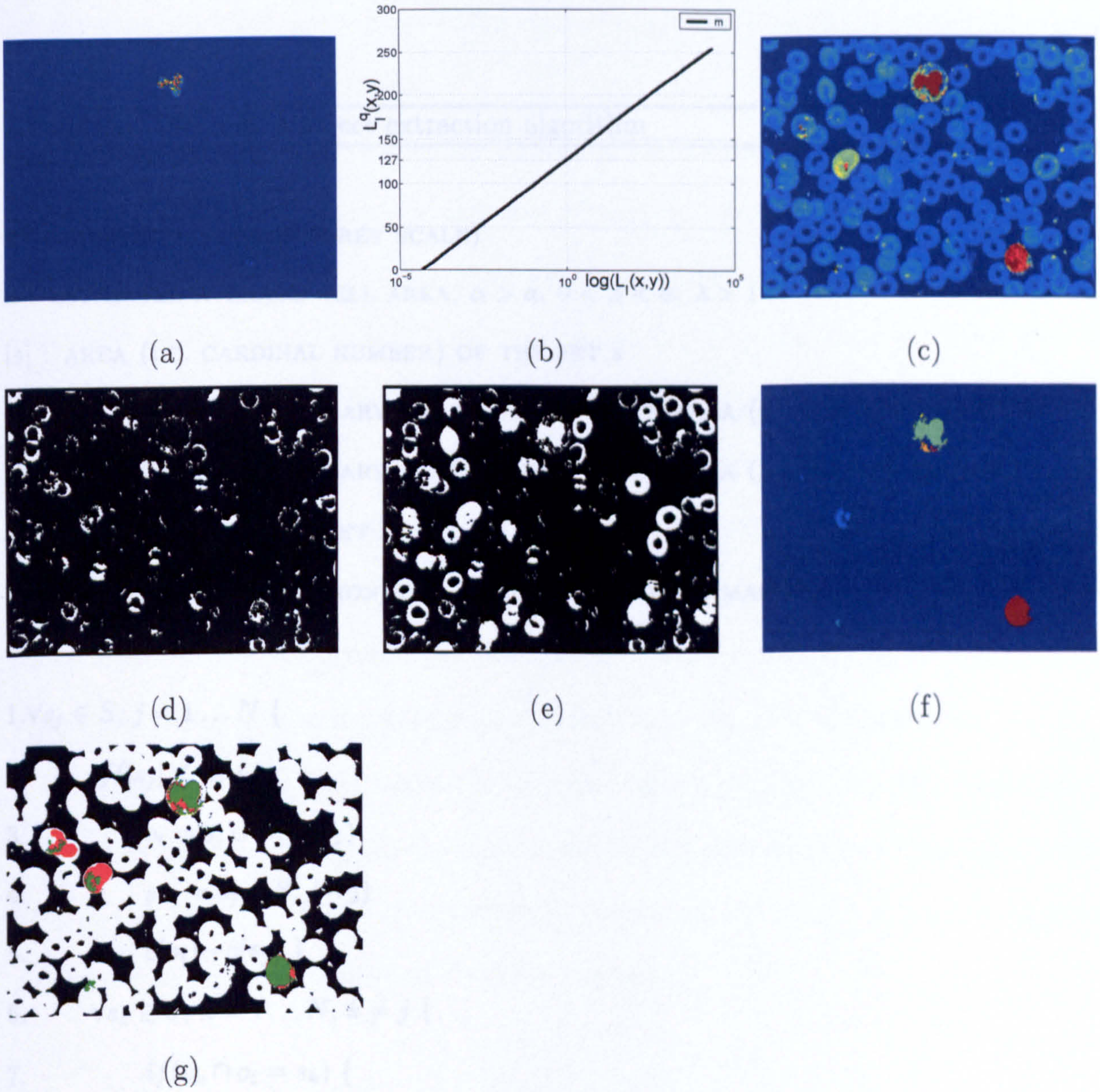


Figure 6.5: Stained object extraction procedure: (a) real valued likelihood image $L_I(x, y)$, (b) the quantisation transfer function, (c) quantised $L_I^q(x, y)$ (red/blue regions correspond to higher/lower values), (d) area top-hat by β, (e) area top-hat by α, (f) input labelled stained pixel groups, (g) extracted stained objects.

Algorithm 3 Stained object extraction algorithm

L_I^q : LIKELIHOOD IMAGE (GREY SCALE)

a : ESTIMATED AVERAGE CELL AREA, $\alpha > a$, $0 < \beta < a$, $\lambda > 1$

$|s|$: AREA (I.E. CARDINAL NUMBER) OF THE SET s

$W_1 = WTH^{\alpha}(L_I^q) > 0$: BINARY THRESHOLDED WHITE AREA (α) TOP-HAT OF THE L_I^q

$W_2 = WTH^{\beta}(L_I^q) > 0$: BINARY THRESHOLDED WHITE AREA (β) TOP-HAT OF THE L_I^q

$c(s)$: THE LABEL OF THE SET s

$R_f^{\delta}(g)$: THE RECONSTRUCTION BY DILATION OF (g) WITH MARKER f

1.$\forall s_j \in S,\ j = 1 \ldots N$ {

2. $if(|s_j| \leq a)$ {

3. $o_i = s_j \cup R_{s_j}^{\delta}(W_1)$

4. $se_j = s_j \cup R_{s_j}^{\delta}(W_2)$

5. $c(o_i) \leftarrow c(s_j)$ }

6. $\forall s_k \in S,\ k = 1 \ldots N,\ k \neq j$ {

7. $if(s_k \cap o_i = s_k)$ {

8. $c(s_k) \leftarrow c(o_i) \quad S \leftarrow S - s_k$ } }

9. $i \leftarrow i + 1$ }

Thus, the average cell area is regarded as an upper limit for the small stained object. In Line 4 an enhanced version of s_j is calculated in order to merge close stained pixel groups within the cell region. In Lines 6 and 8, the algorithm searches for the stained pixel groups (s_k) that are subsets of the stained object o_i; then changes their label and excludes from further processing.

The large stained objects can be WBCs, artefacts and some mature forms of P. Vivax and P. Ovale species. In the case of mature parasites the stained object is usually equivalent to the stained pixel group. Therefore, further processing is not necessary. In the case of a WBC they can be enhanced by the same approach that is followed for small stained objects; however there is no upper limit that can be relied on. Moreover, it is not guaranteed that the stained pixel group is a partially found WBC: it can be due to an artefact which can be in a random location (e.g. on top of RBCs). Hence, the processing of large stained objects to enhance WBC extraction is possible but at the cost of enlarging the artefacts which can be on top of RBCs. This was implemented but excluded later due to negative effects on the artefacts.

Figures 6.5(d) and 6.5(e) show the thresholded area top-hats W_1, W_2 respectively. In Figure 6.5(f) and 6.5(g) the labelled image of the stained pixel groups and corresponding output can be seen, respectively. Note that in the output some stained pixel groups are enlarged, and some of them are merged.

6.5 Experimental Results

The experiments and results are explained in two parts. First the evaluation experiments for the Bayesian stained pixel detection method are explained. Later, the advantages and disadvantages of the stained object extraction algorithm are discussed with some examples.

6.5.1 Stained/Non-Stained Pixel Classification

The basics of the classifier and the computation of the histograms are explained in Section 6.2. In order to realise the Bayesian pixel classifier, the histograms of stained (w_s) and non-stained (w_{ns}) classes must be computed and the likelihood threshold θ must be found. Additionally, since the histograms are estimates of class conditional density functions, the optimal histogram size Q for our finite number of samples has to be determined. Hence, two separate sets of labelled (i.e. supervised) samples are required: a training set and a validation set. The training set T_a is to compute the histograms; the T_θ validation set is to estimate the parameters θ and Q. Finally, a separate set of labelled samples T_t is required to assess the generalisation performance of the trained classifier with estimated parameters.

The preparation of the stained pixel ground-truth data was explained in Section 3.2.1. By using a simple program written for the purpose the stained pixel components in 200 images were labelled by manual thresholding. The stained pixel data were represented by binary masks. The images were then divided into three separate sets: training (T_a), validation (T_θ), and testing (T_t) that contained (100/-211635/ 28080650), (35/86820/9247932), (65/132613/18159323) images/stained/non-stained pixels, respectively. The images were pre-processed with the colour normalisation method presented in Chapter 5. However, in order to compare the effects of this process images were also used without colour normalisation.

Performance Measures: The distribution of the samples between the two classes was unbalanced (i.e. skewed data). There were more than 9 million non-stained pixels for only 86820 stained samples in the validation set T_θ. Therefore, average classification accuracy (6.13), which is the ratio of the total number of the correct classifications to the total number of samples, was not very informative. For example, we can obtain 99% average accuracy by simply assigning all the pixels to non-stained pixel class. Therefore the errors or the accuracies have to be analysed separately. Let S_i, G_i, and

G_i^c show the detected pixels by the classifier, the manually marked binary image of the stained pixels, and its complement for the ith image (I_i) in the set (T_θ or T_i), respectively; the true and false detection rates were calculated as follows:

$$T.D = \frac{\sum_i \sum_{I_i} (S_i \cap G_i)}{\sum_i \sum_{I_i} G_i}, \qquad F.D = \frac{\sum_i \sum_{I_i} (S_i \cap G_i^c)}{\sum_i \sum_{I_i} G_i^c} \tag{6.12}$$

The average accuracy measure can be calculated as follows:

$$Acc = \frac{T.D + c(1 - F.D)}{1 + c} \qquad c = \frac{\sum_i \sum_{I_i} G_i^c}{\sum_i \sum_{I_i} G_i} \tag{6.13}$$

A true detection refers to a stained pixel that is correctly assigned to the stained class; whereas a false detection refers to a non-stained pixel that is incorrectly assigned to the stained class. $T.D$ and $F.D$ reflects the probability of these within the sample set.

The skew insensitive accuracy measure which can be an alternative to Acc can be calculated as follows [49]:

$$Acc^* = \frac{T.D + (1 - F.D)}{2} \tag{6.14}$$

For T_θ, c was 9247932/86820 = 106.5. By this ratio, the Acc^* measure is in favour of the stained class.

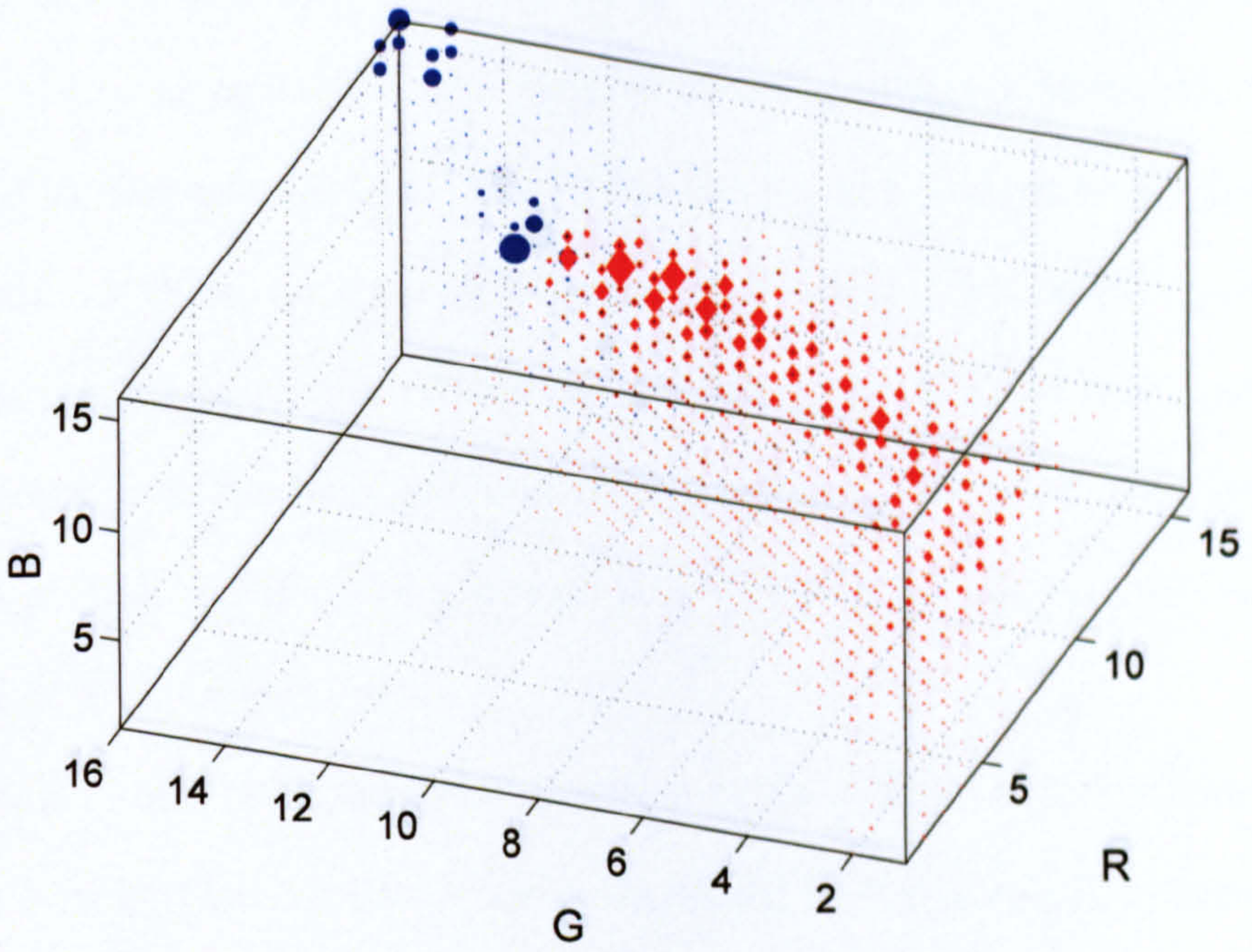
The histograms: The histograms are the count of occurrences of $[r, g, b]$ triples separately for the stained and non-stained classes. However, since they are normalised respectively by the total count of samples in each class they reflect the conditional probabilities which are 3 dimensional real valued matrices. In order to observe the distribution of the two classes and the colour normalisation effects, the scatter plots of the histograms were produced. Figure 6.6(a) and 6.6(b) show 3-d scatter plots of the histograms for the stained and non-stained classes obtained from the samples in T_a for two example histograms. Although, the colour channels in the input images are normally quantised to 256 discrete levels, the scatter plotted histograms are in 16 and 32 discrete bins for each channel. This is discussed in more detail in the next paragraph. The scatter plots are superimposed: the blue circles represent the non-stained; the red

diamonds represents the stained class. The size of the marker (i.e. circle, diamond) shows the amplitude. It can be observed that the 32^3 histogram (Figure 6.6(b)) is more pronounced. In both cases, it can be observed that the colour normalisation separates two clear, clean peaks (i.e. cells, plasma) of non-stained pixels. Therefore, the spreading of non-stained pixels is concentrated on the two peaks. The same plots are produced for the same pixel samples (Figure 6.7(a) and Figure 6.7(b)); however without colour normalisation applied to the images. In the non-normalised case, the spread in stained pixels is pronounced; the non-stained pixels are not concentrated into two peaks (i.e. cells, plasma); stained and non-stained pixels groups are more intricate.

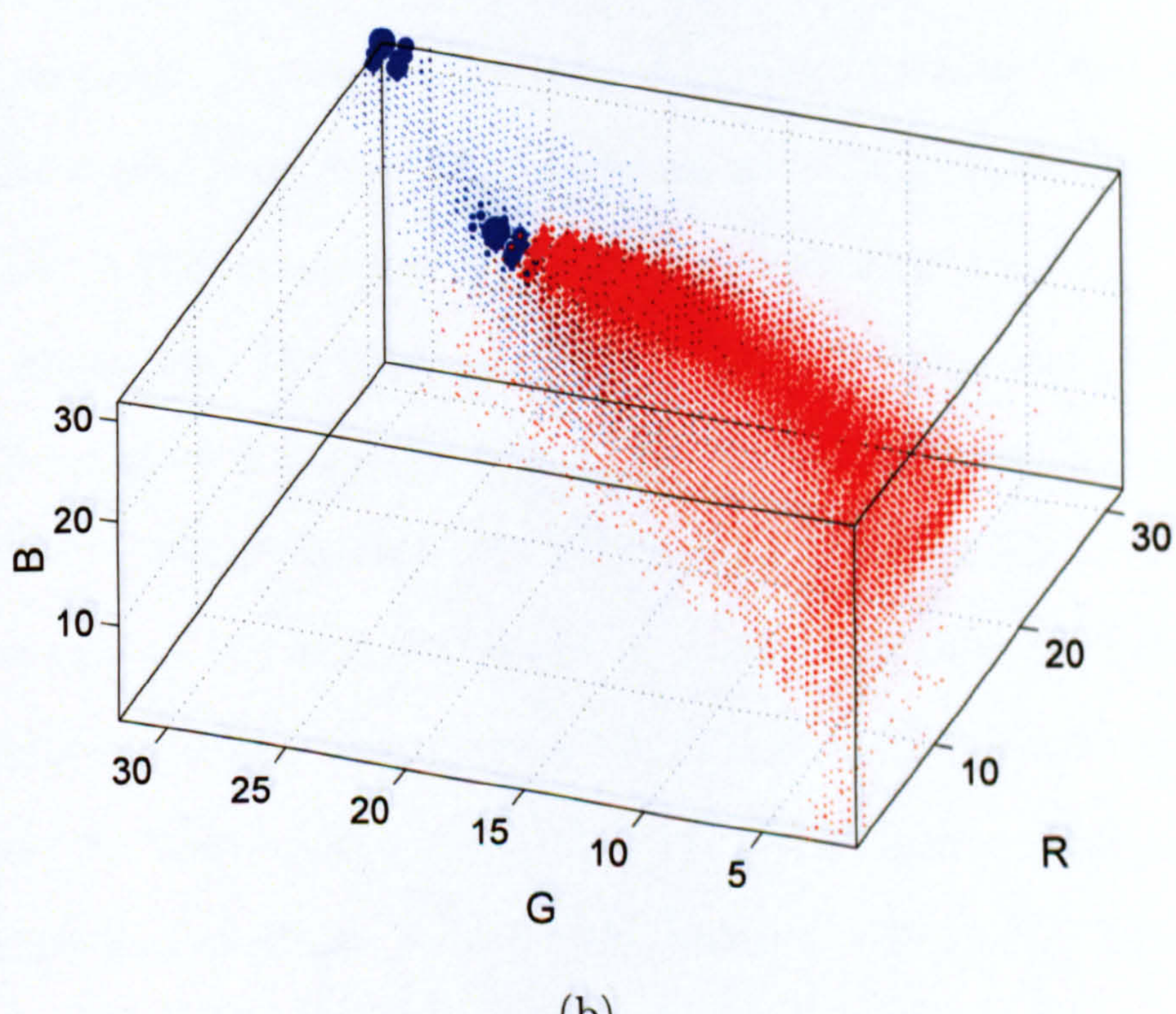
The histogram size Q and likelihood threshold θ: It is desirable to reduce the number of levels (i.e. re-quantise) so the histogram size is manageable and more efficient for the density estimation. The Q value was determined by comparing the ROC curves of the different histogram sizes.

Figure 6.8(a) and 6.8(b) show different ROC curves (for T_θ) of different sizes of histograms (quantisation levels :$256, 128, 64, 32, 16, 8$). See Section 6.3 for a brief explanation of ROC analysis. For the colour normalised case, the ROC plots for the different histogram sizes are plotted in Figure 6.8(a). Since the class distributions unequal (i.e. positive/negative ratio 106.5), we are interested in a limited interval of this curve. Figure 6.8(b) shows 0-1% interval of the original ROC curves. Unfortunately, the curve for 8^3 size histogram was crossing the other curves near $F.D = 0.1$ value but was below $16^3, 32^3, 64^3$ curves for the rest of the $0 - 1\%$ interval. Ignoring this single peak value, it can be concluded that most efficient histogram size (for T_a) is 16^3 with the largest area under the curve. The θ value is determined on the ROC curve points by the value that produces the maximum accuracy.

However, we have plotted both the accuracy (Acc (6.13)) and skew insensitive accuracy (Acc^* (6.14)) values versus increasing θ. Then we determined the maximum θ indices, for each case, separately (see Figure 6.9(a) and Figure 6.9(b)). For the 16^3 histogram size, threshold values $\theta = 1$ and $\theta = 177.8$ provided the peak values for Acc^*

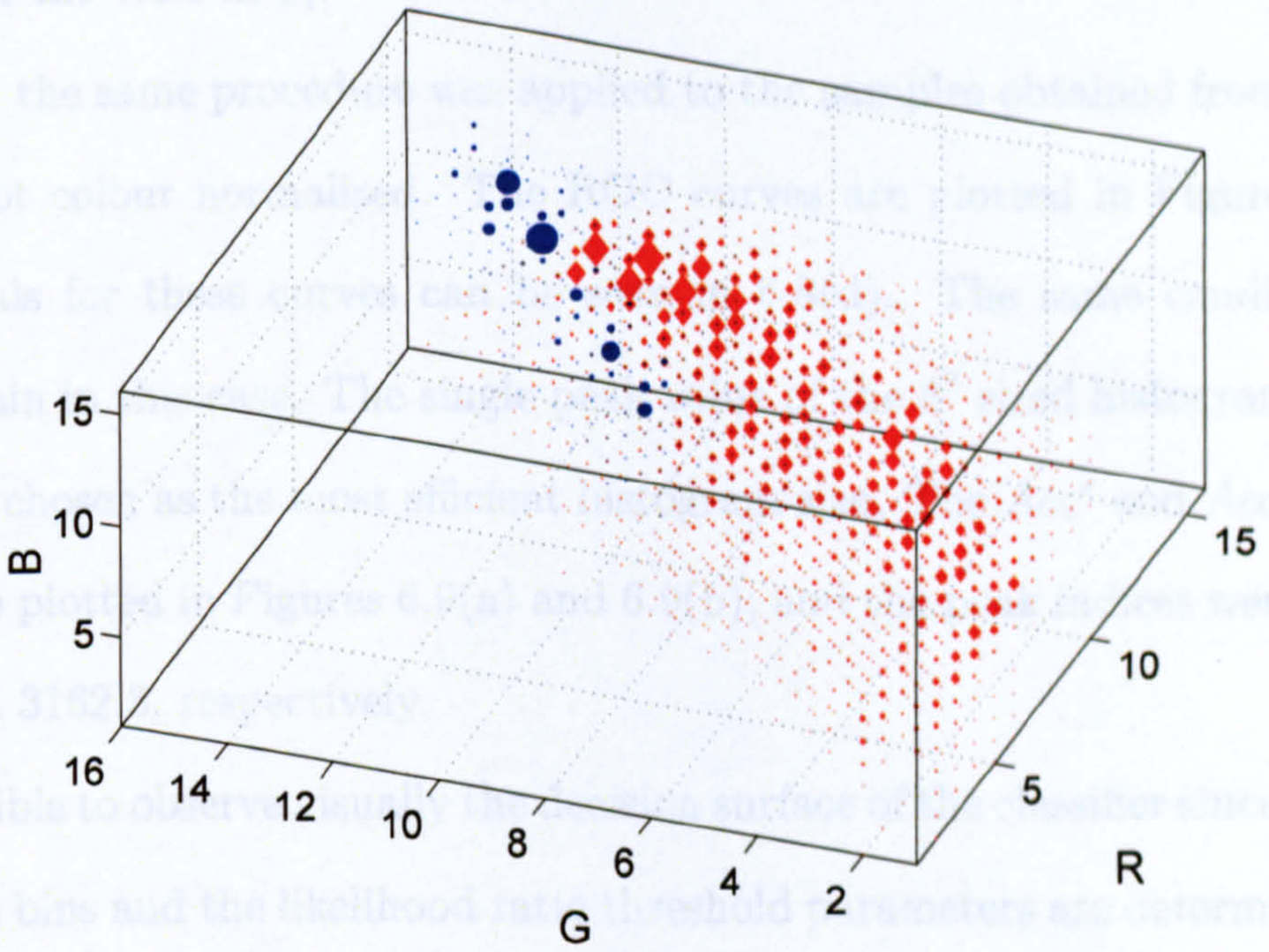


(a)

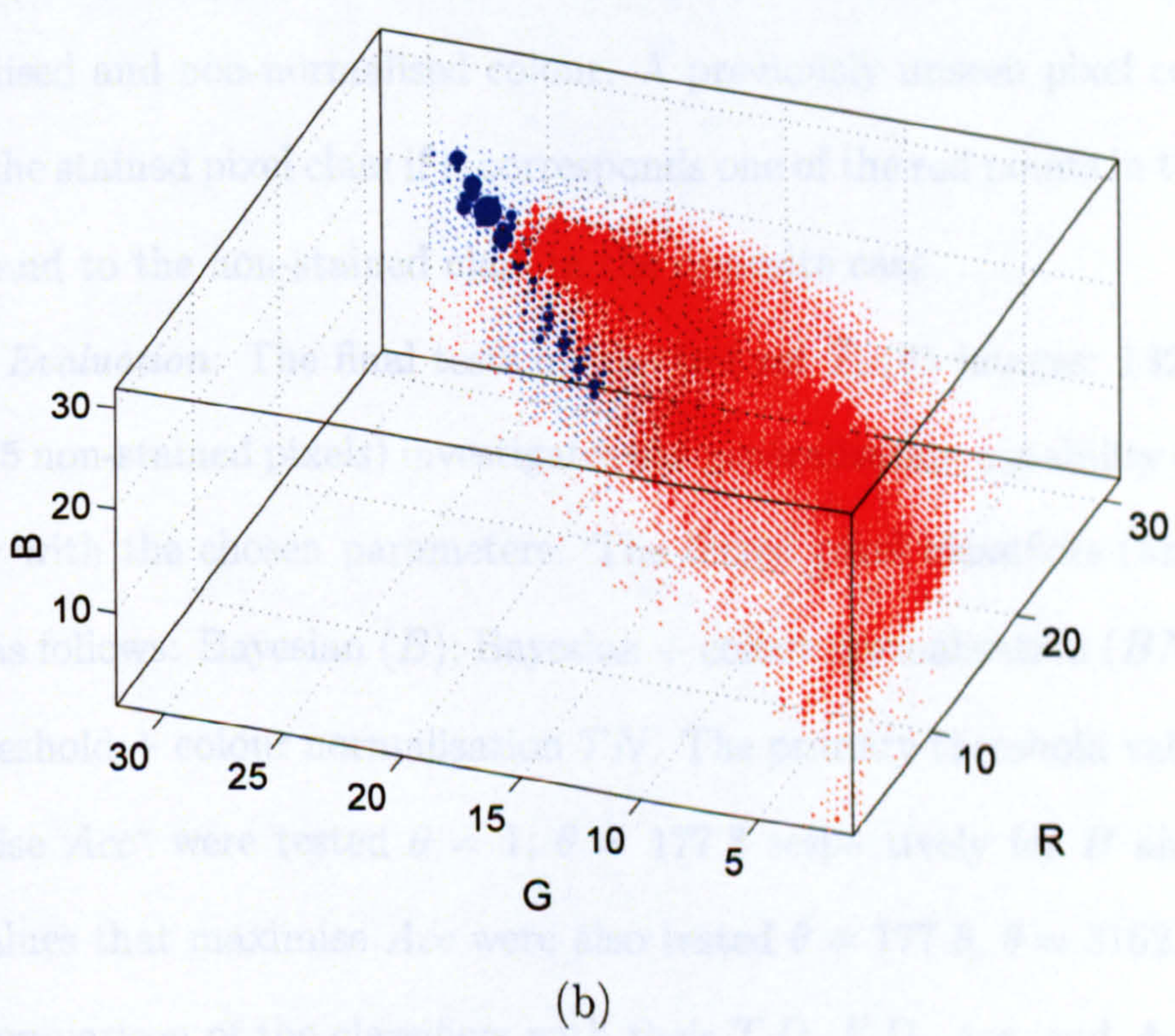


(b)

Figure 6.6: Scatter plots of the histograms (with colour normalisation): (a) 16^3 bins, (b) 32^3 bins. The red (diamonds) and blue (circles) denote stained and non-stained classes, respectively. Marker sizes show relative amplitude.



(a)



(b)

Figure 6.7: Scatter plots of the histograms (without colour normalisation): (a) 16^3 bins, (b) 32^3 bins. The red (diamonds) and blue (circles) denote stained and non-stained classes, respectively. Marker sizes show relative amplitude.

and *Acc* measures, respectively. These θ values together with the 16^3 bin histogram were used for the tests in T_t.

However, the same procedure was applied to the samples obtained from the images that were not colour normalised. The ROC curves are plotted in Figure 6.8(c) and 0-1% intervals for these curves can be seen in 6.8(d). The same condition can be observed again in this case. The single peak value of the 8^3 sized histogram is ignored, and 16^3 was chosen as the most efficient histogram size. The Acc^* and Acc values with respect θ are plotted in Figures 6.9(a) and 6.9(b); and the peak indices were calculated as 177.8 and 3162.3, respectively.

It is possible to observe visually the decision surface of the classifier since the number of histogram bins and the likelihood ratio threshold parameters are determined. Figure 6.10(a) and 6.10(b) plots the scatter plot of the pixel triples that have the likelihood value above (red points) and below (blue triangles) the threshold (θ), respectively for both normalised and non-normalised colour. A previously unseen pixel colour will be assigned to the stained pixel class if it corresponds one of the red points in the quantised RGB space and to the non-stained class in the opposite case.

Test Set Evaluation: The final tests on the test set T_t (65 images: 132613 stained, and 17901275 non-stained pixels) investigates the generalisation capability of the Bayesian classifier with the chosen parameters. The tested pixel classifiers (and configurations) were as follows: Bayesian (B); Bayesian + colour normalisation (BN); threshold (T) [35]; threshold + colour normalisation TN. The primary threshold values obtained that maximise Acc^* were tested $\theta = 1$, $\theta = 177.8$ respectively for B and BN. The threshold values that maximise Acc were also tested $\theta = 177.8$, $\theta = 3162.3$. Table 6.1 presents a comparison of the classifiers with their $T.D$, $F.D$, Acc, and Acc^* rates.

Colour Normalisation: It can be seen that in both set of thresholds (Acc and Acc^*) the colour normalisation reduces the false detection rate of the Bayesian classifier and increases the true detection rate. If the number of samples for the non-stained class is considered, the difference of the false detection rates is quite significant. The difference

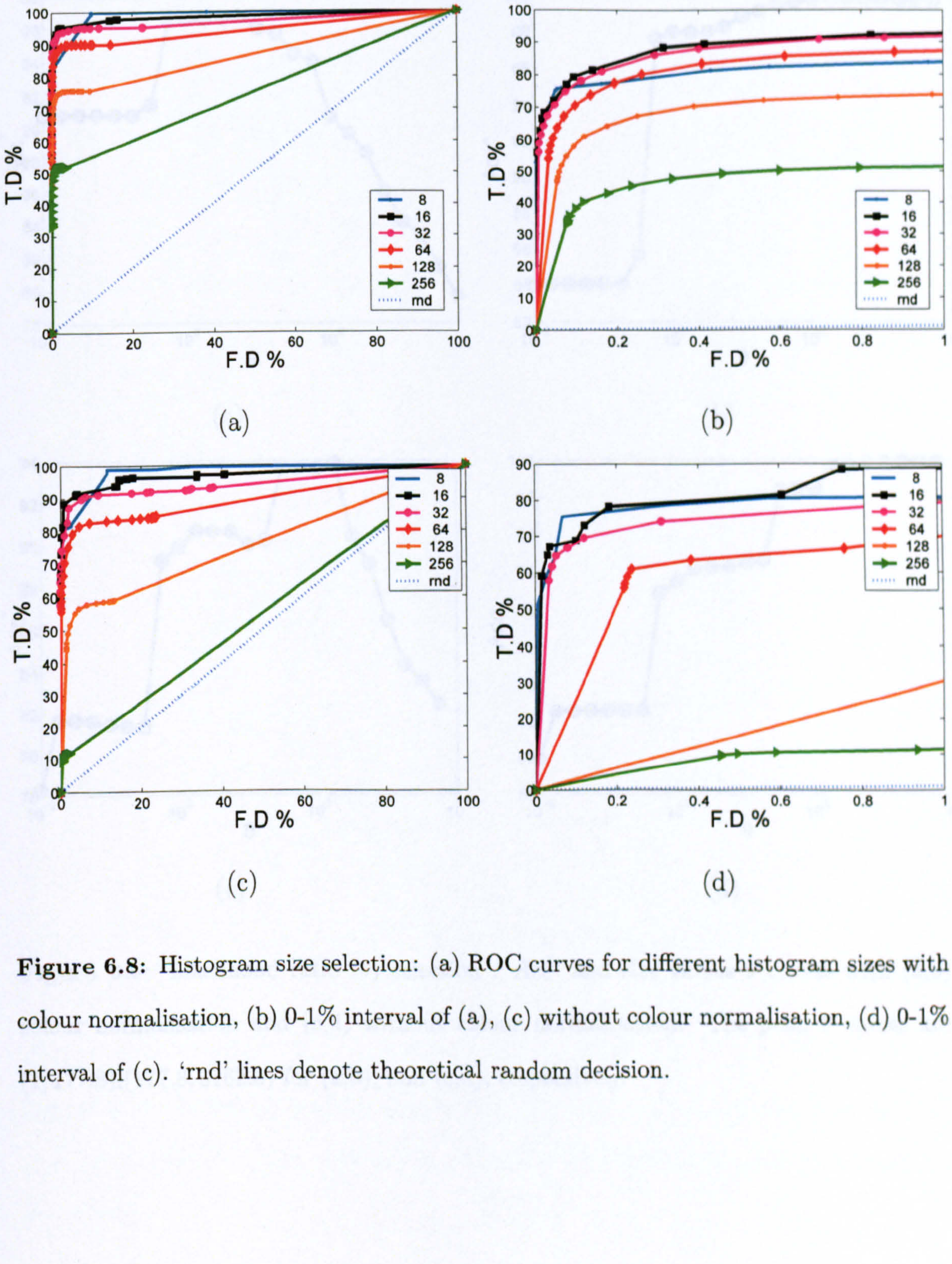


Figure 6.8: Histogram size selection: (a) ROC curves for different histogram sizes with colour normalisation, (b) 0-1% interval of (a), (c) without colour normalisation, (d) 0-1% interval of (c). 'rnd' lines denote theoretical random decision.

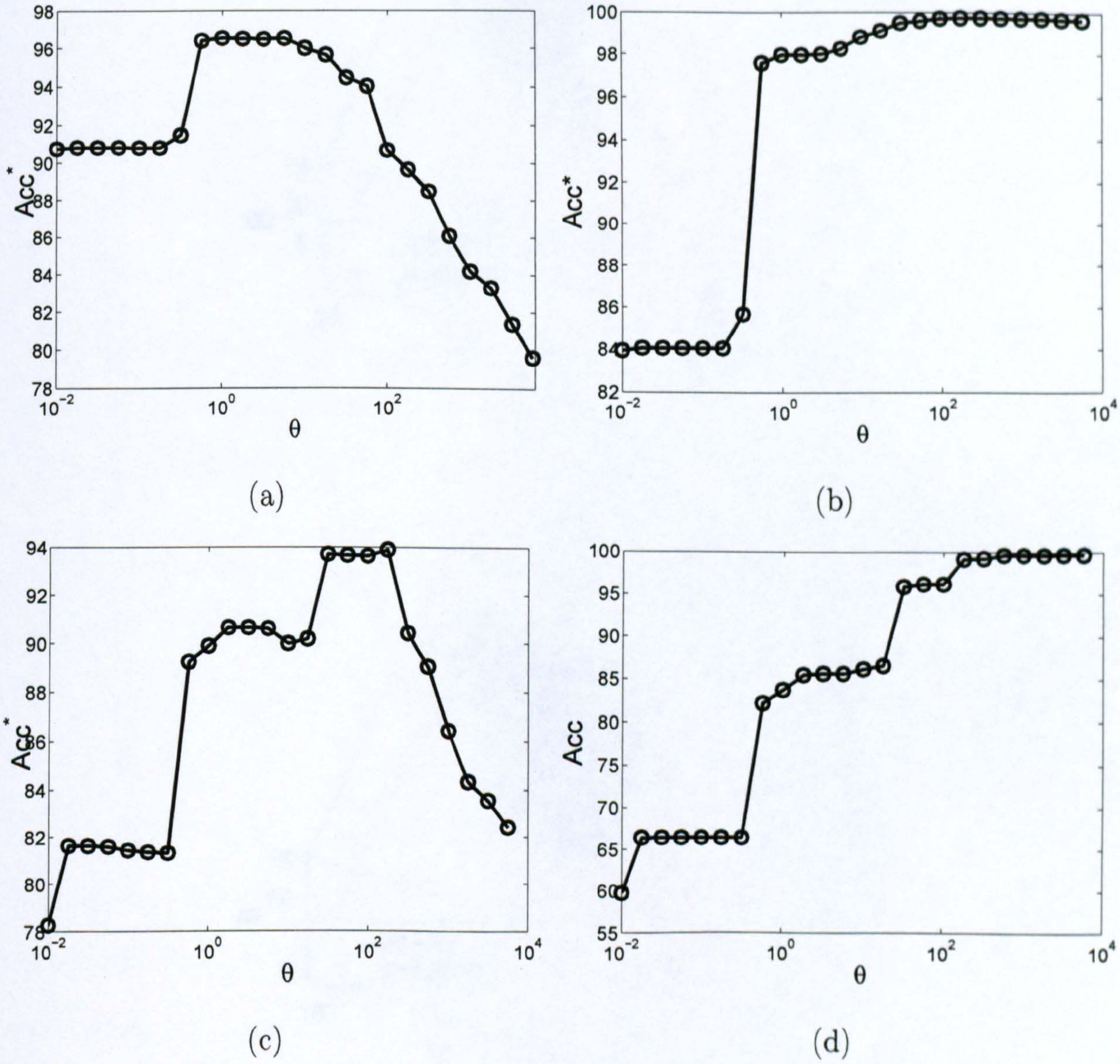
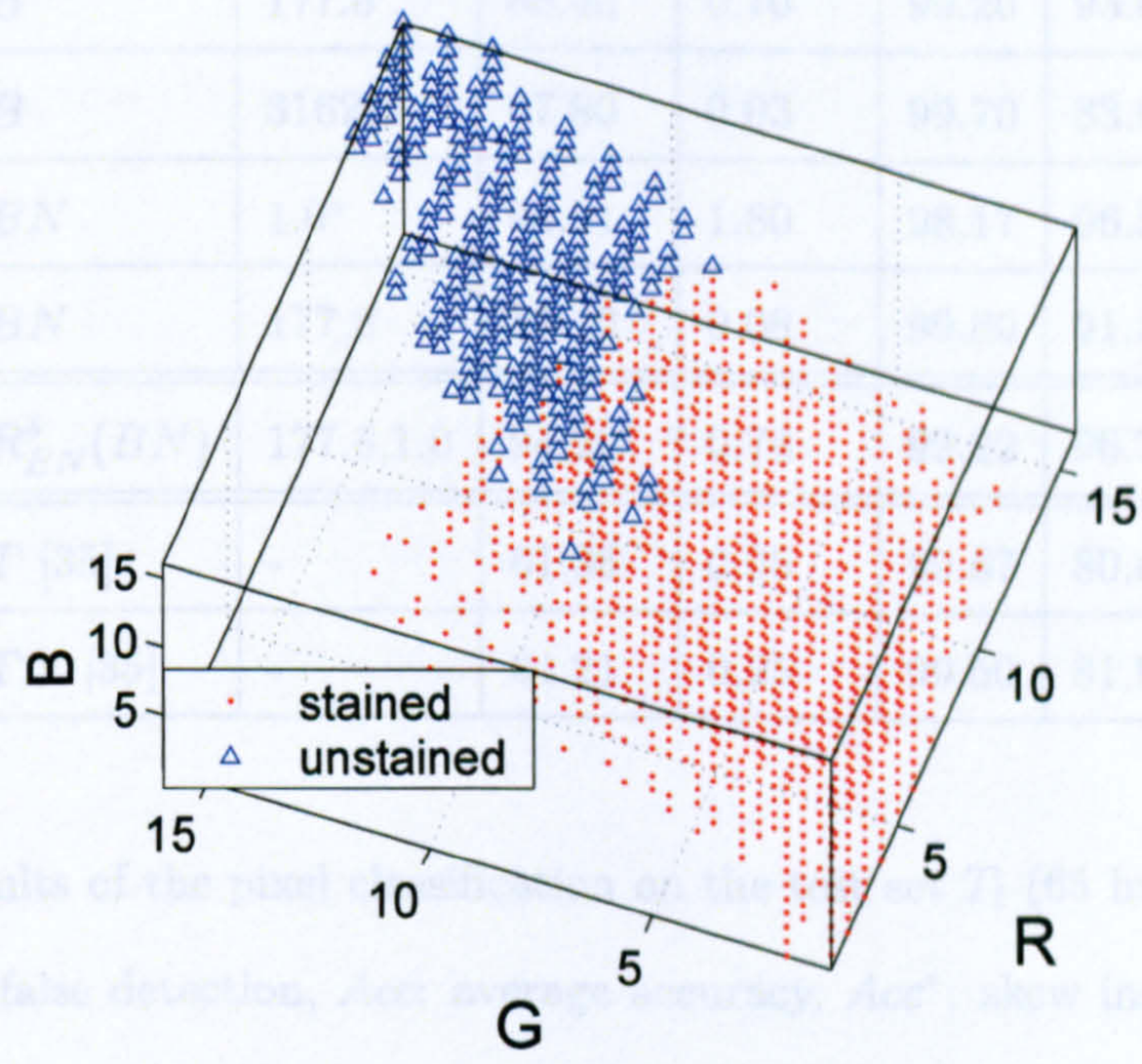
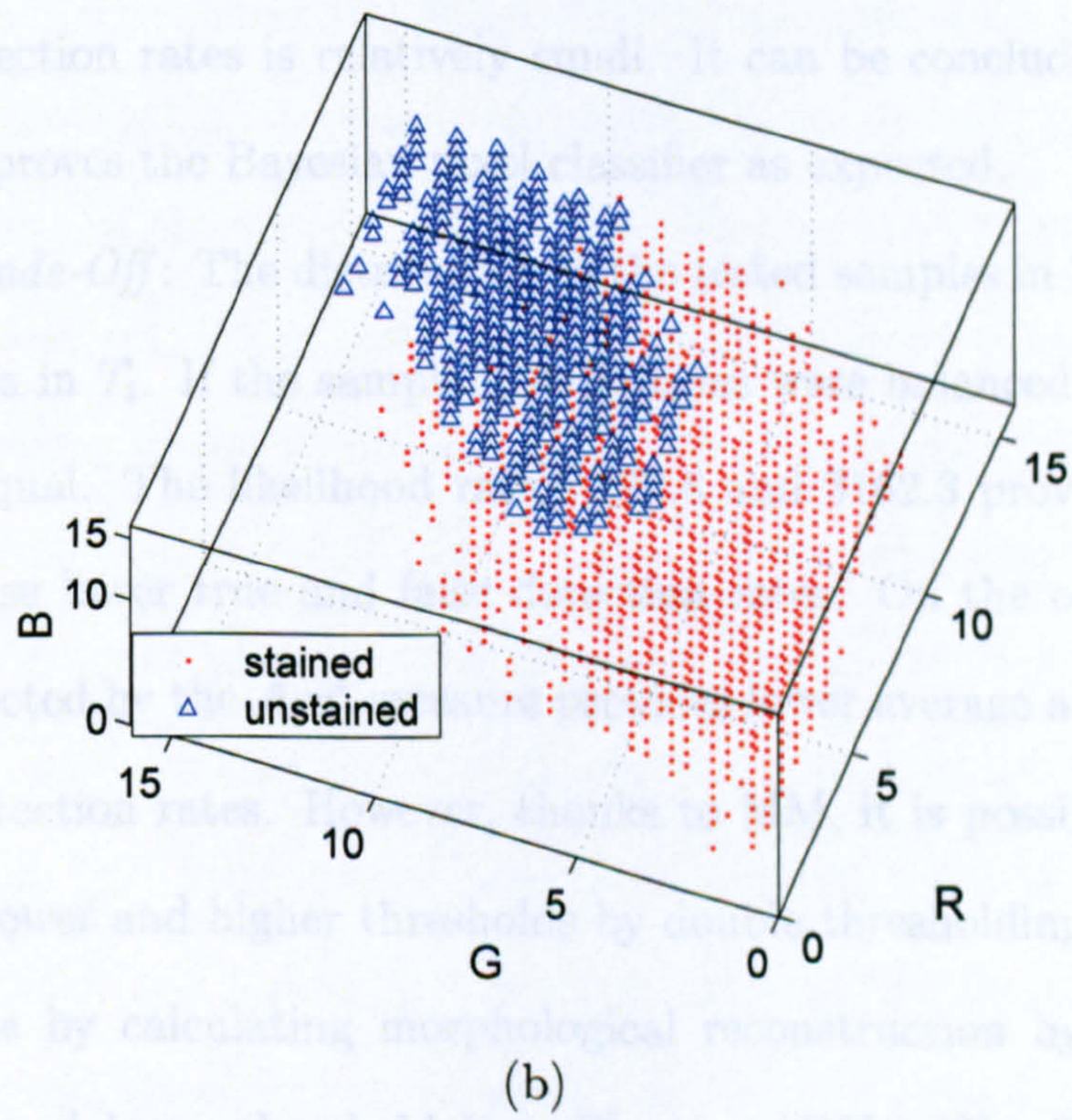


Figure 6.9: Likelihood ratio (θ) selection : Acc^* and Acc versus θ curves with (a,b) colour normalisation and (c,d) without colour normalisation. The peak θ values are $(1, 177.8)$,$(177.8, 3162.3)$ for (a,b), and (c,d), respectively.



(a)



(b)

Figure 6.10: Scatter plots of the likelihood: (a) $L_I(x, y) > \theta$ red (points), $L_I(x, y) <= \theta$ blue (circles) for ($\theta = 1.0$) with colour normalisation (b) without colour normalisation ($\theta = 177.8$).

Method	θ	$T.D.\%$	$F.D.\%$	Acc	Acc^*
B	177.8*	88.40	0.70	99.20	93.88
B	3162.3	67.80	0.03	99.70	83.91
BN	1.0*	94.91	1.80	98.17	96.55
BN	177.8	83.12	0.08	99.80	91.52
$R^{\delta}_{BN}(BN)$	177.8,1.0	94.20	0.74	99.22	96.72
T [35]	-	61.35	0.05	99.67	80.65
TN [35]	-	64.21	0.25	99.50	81.98

Table 6.1: Results of the pixel classification on the test set T_t (65 images). $T.D$: true detection, $F.D$: false detection, Acc: average accuracy, Acc^*: skew insensitive accuracy rates. The values are rounded to second decimal.

between true detection rates is relatively small. It can be concluded that the colour normalisation improves the Bayesian pixel classifier as expected.

$Acc-Acc^*$ Trade-Off: The distribution of the tested samples in T_t was not uniform among two classes in T_t. If the sample distributions were balanced the Acc and Acc^* rates would be equal. The likelihood ratios 177.8 and 3162.3 provide higher average accuracy but cause lower true and false detection rates. On the contrary, the ratios 1.0 and 177.8 selected by the Acc^* measure provides lower average accuracy but higher true and false detection rates. However, thanks to MM, it is possible to make use of positive sides of lower and higher thresholds by double thresholding (see Section 4.3). This can be done by calculating morphological reconstruction by dilation between higher threshold and lower threshold (i.e. $R^{\delta}_{BN(177.8)}(BN(1.0))$. Thus, by using the higher threshold image as the markers for the reconstruction of the lower threshold image, some of the false detections may be reduced. This was tested on T_t for colour normalised images and achieved considerably better results as shown in Table 6.1.

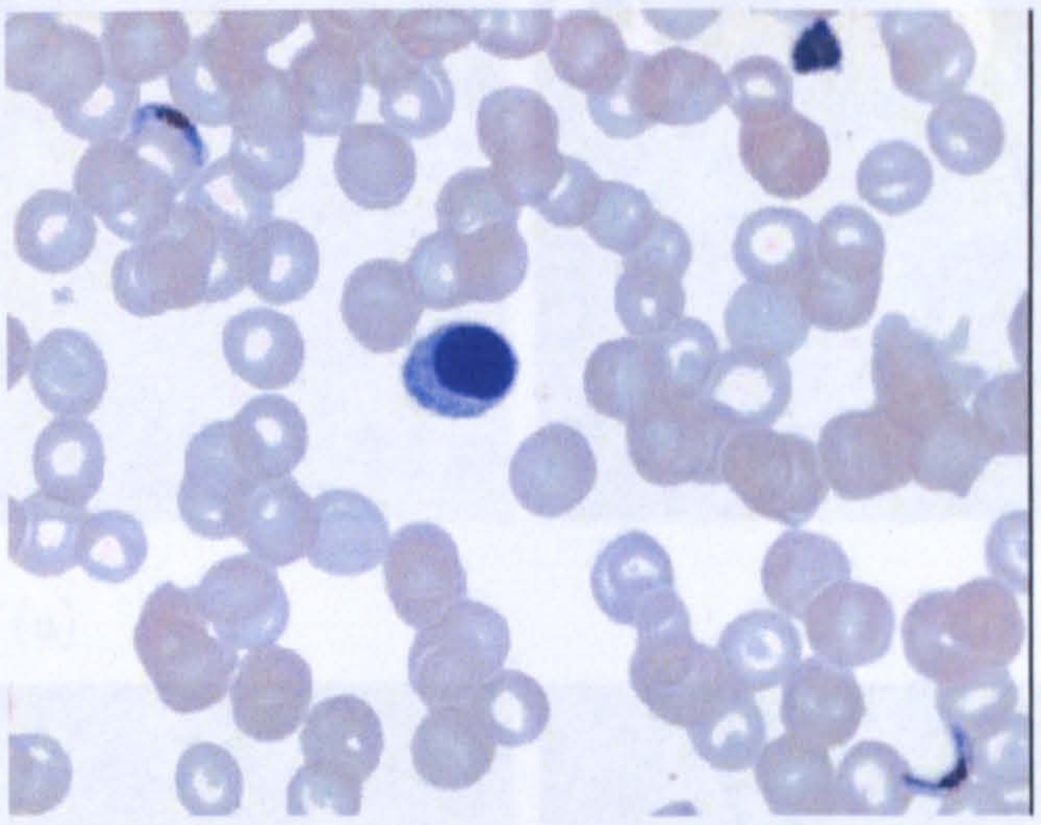
The threshold method [35] produced significantly lower true detection rate than the

Bayesian classifiers. However, the false detection rate was quite low which suggests a moderate performance. The effects of colour normalisation (TN) were opposite: both true and false detection rates were increased. The increase in true detection rate was not significant.

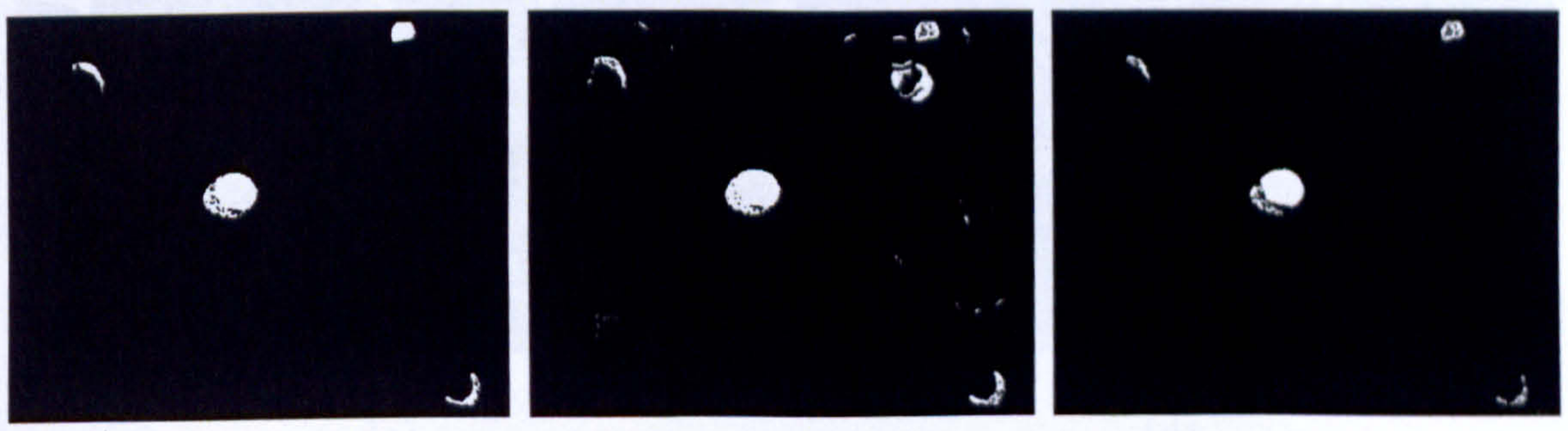
Figure 6.11 provides a visual comparison of the different configurations and threshold settings for the Bayesian classifier together with the results of the threshold method. It can be observed that the Bayesian classifier with double thresholding produces the closest output to the ground truth than the single threshold classifiers and the threshold method.

Discussion: The decision rule of the Bayesian pixel classifier is learned from previously seen samples and so the decision is regardless of the new input image: it uses the supervised training samples to divide the RGB colour space. In general, this division is not required to be linear and there is no ordering required between two classes (i.e. stained and non-stained sample grey levels). However, the samples used in the training, validation and testing of the classifier were produced by manual thresholding operations by observing the grey level histogram and the input-thresholded images (see Section 3.2.2). Hence, all the samples that are marked are ordered in a particular input image: the stained pixel samples always have smaller grey value than non-stained pixel samples. Thus, the stained pixel ground-truth information put a bias on the tests. This contributes to the higher false detection rates. In other words, manual thresholding on the grey level images (and so the ground-truth information) assumes that the stained pixels are darker (grey level value) than the non-stained pixels in an input image; however, Bayesian pixel classifier decides the class of the pixel according to its RGB colour vector. Hence, it can assign some colour pixels to the stained pixel class even they are not darker than non-stained pixels and (and were not marked in the ground truth) in an input image.

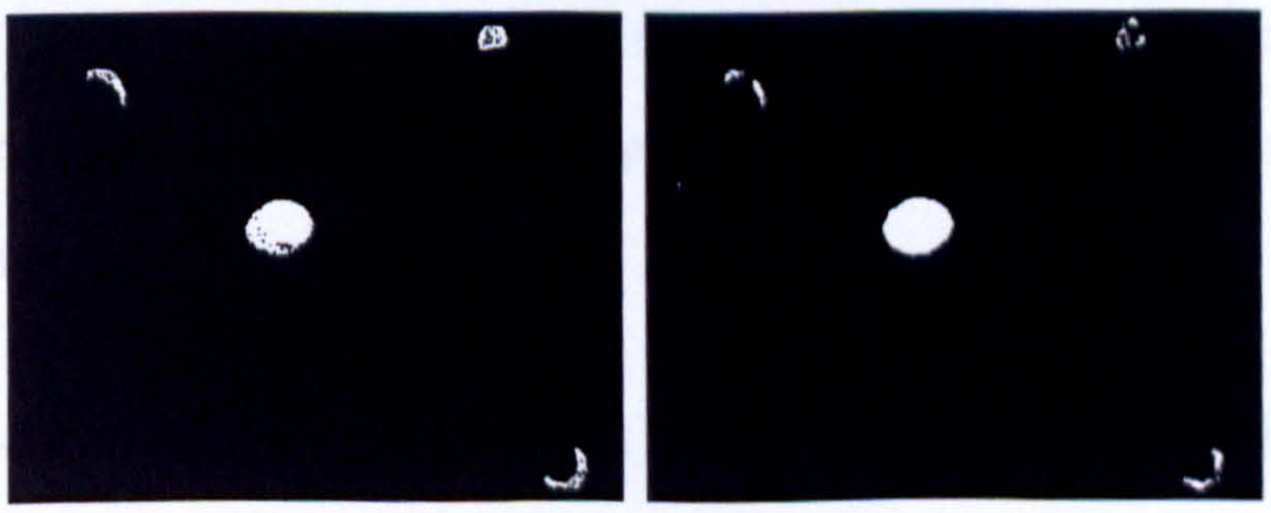
This is issue is demonstrated in Figure 6.12 with an input example. It can be seen that not all marked pixels in the output of the Bayesian pixel classifier are not darker



(a)



(b) (c) (d)



(e) (f)

Figure 6.11: Comparison of the pixel detection results: (a) colour normalised input image, (b) manual threshold, (c), (d) Bayesian pixel classifier with thresholds 1.0 and 177.8, respectively, (e) Bayesian pixel classifier with double threshold $R^{\delta}_{BN(177.8)}(BN(1.0))$, (f) the threshold method [35].

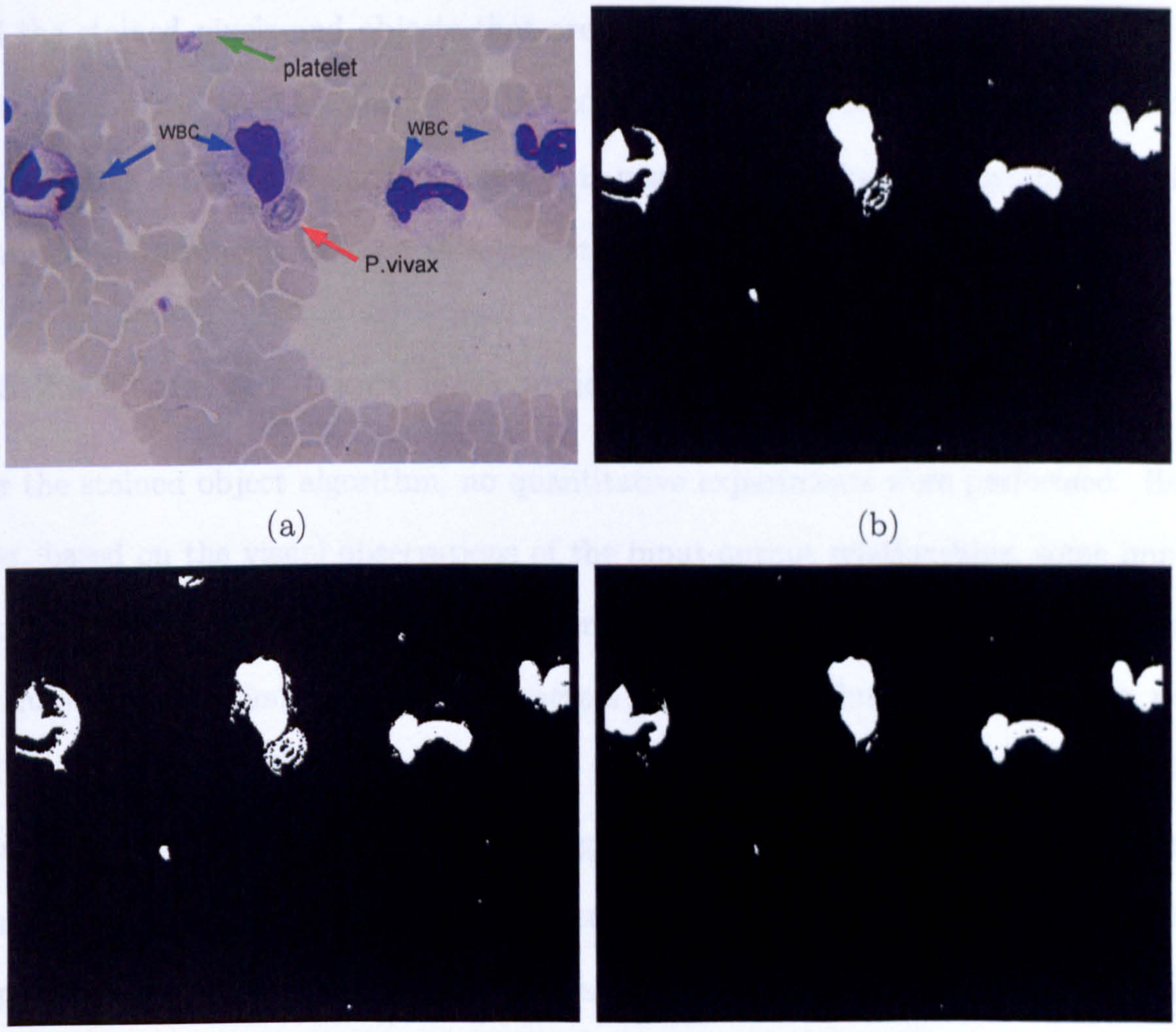


Figure 6.12: Comparison of the pixel detection results: (a) colour normalised input image, (b) manual threshold, (c) Bayesian pixel classifier with double threshold, (d) the threshold method [35].

than the non-stained pixels (Figure 6.12(c)). The manual marking (Figure 6.12(b) is limited with clearly darker pixels (e.g. does not include top-centre platelet indicated by green arrow). The threshold method detects only some of the darker/saturated pixels (nearly missing the parasite indicated by red arrow).

For CDM, missing a single stained pixel may not be critical because the parasites appear in our test image resolutions with more pixels (proportional to the maturity of the parasite). However, missing a group of pixels (connected) may change the diagnosis result of the CDM system: in the case of a parasite, it reduces sensitivity of the system.

All the stained pixels and objects that are marked by them are further classified to detect the parasites (Chapter 8); on the contrary a missed stained group is excluded from further analysis. Therefore, even it is possible to utilise the Bayesian classifier with higher thresholds to lower false detection rates, this may not be preferable.

6.5.2 Stained Object Extraction

For the stained object algorithm, no quantitative experiments were performed. However, based on the visual observations of the input-output relationships, some important aspects are discussed here. The average cell area a is determined by the area granulometry (Section 4.2). The parameters for the algorithm (Algorithm 3 on page 133) were heuristically determined; set to the values $\alpha = 1.1a$, $\beta = 0.25a$. The first parameter α is used to detect the RBC regions that contain the stained pixel groups. The second parameter β is used to enhance the small stained pixel groups and merge if possible. The choice of values are discussed in the next two paragraphs with some example input-output pairs which are shown in Figures 6.13-6.16.

- α: If the input image is from a field which is not densely populated, the area top-hat calculation by α which has a slightly bigger value than the input image average cell area (i.e. $\alpha = 1.1a$) extracts the actual cell regions that are separated from other cells and have resolvable boundaries (Figure 6.13). However, if the input image is from a densely populated, concentrated field, the cell boundaries may not be totally resolvable. Hence, the area top-hat and the following reconstruction operation may extract an arbitrary region that is connected to the input stained pixel group. Figure 6.14 demonstrates this with an input-output image pair. It can be observed that some of the cell borders are not resolvable in the input image; hence some of the stained objects' regions are arbitrary aligned. These regions correspond to the domes, of the likelihood image L_I^q (6.11), which have smaller area than the average cell area. If the stained pixel group of an arte-

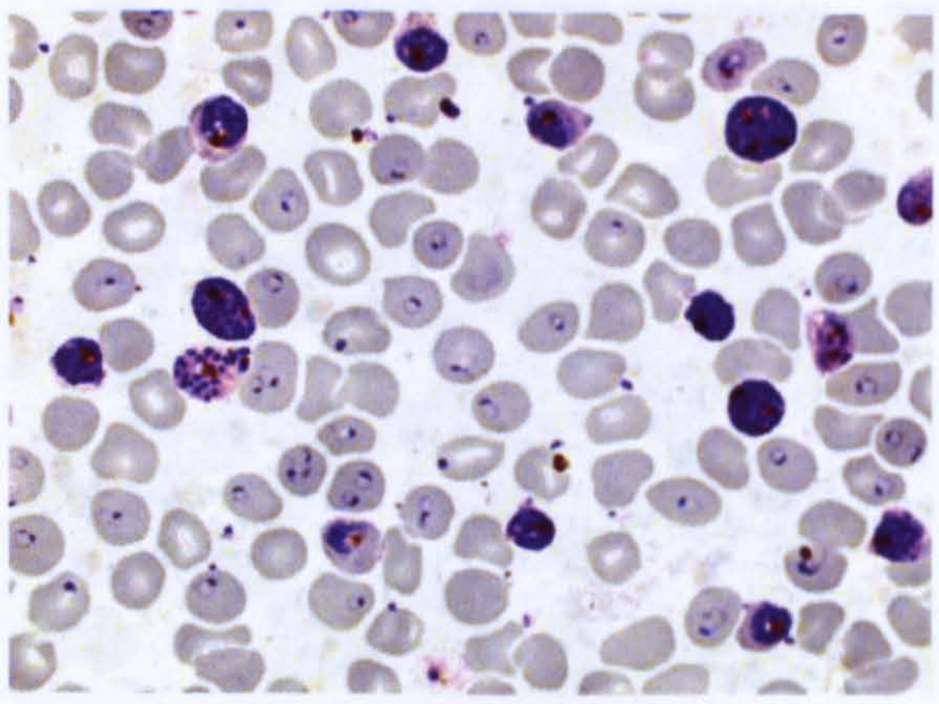
fact or platelet is isolated on the background, the extracted stained object region is usually correct. However, when it touches to (or it is on top of) a RBC the area top-hat propagates through the cells which results in an incorrect region. Since the algorithm processes only the stained pixel groups which are smaller than the average cell, the larger objects are not affected. For example, the WBCs that are located in the centre of the image (Figure 6.16(a)) are left unchanged after the process; however the two *P. Falciparum* (left-top, right-bottom corners) parasites are extracted.

- β: The β parameter controls the enlargement of the partially found stained pixel groups. For example, in a cell infected by the immature ring forms, the Bayesian pixel detection process sometimes detects only the chromatin dots of the parasites (Figures 6.13(b)). By the area top-hat and reconstruction operations (Algorithm 3) the detected stained pixel group(s) is/are enlarged to define all of the parasite region(s) Figure 6.13(c). However, this enlargement is limited by the heuristically chosen value $\beta = 0.25a$. If a stained pixel group is already bigger than the estimated average cell area, it is not affected by this process.

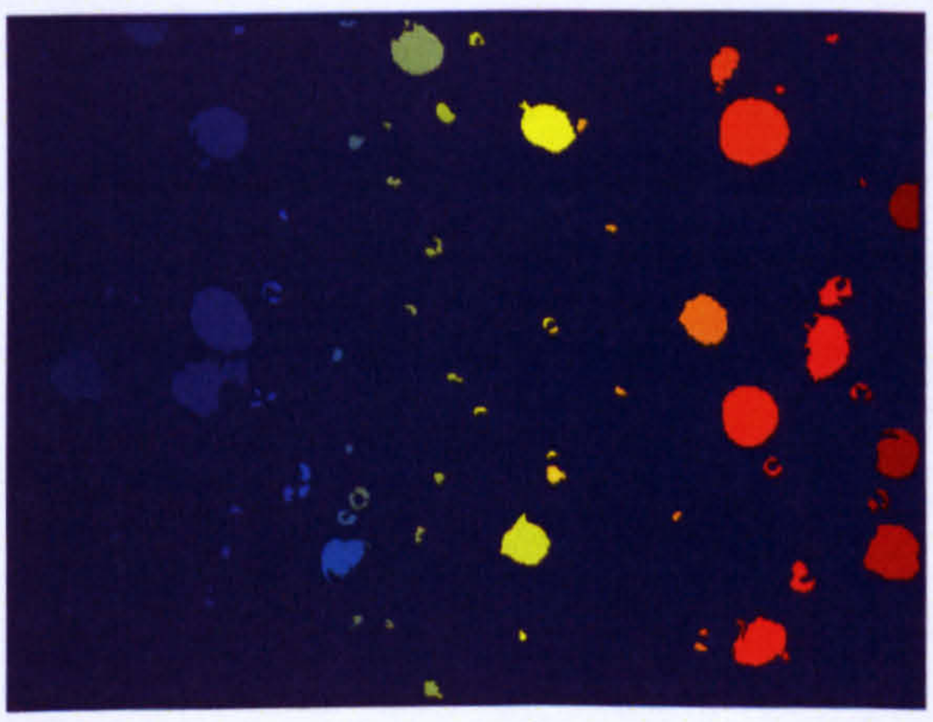
  Alternatively, it was possible to calculate β according to the initial stained pixel area $|s|$ (e.g. $0.25|s|$). However, this was not more efficient since there is no guarantee that the input stained pixel group is partially found and an enlargement may be necessary. On the contrary, by fixing the β value to be $0.25a$ the enlarged region is limited by this value.

6.6 Conclusions

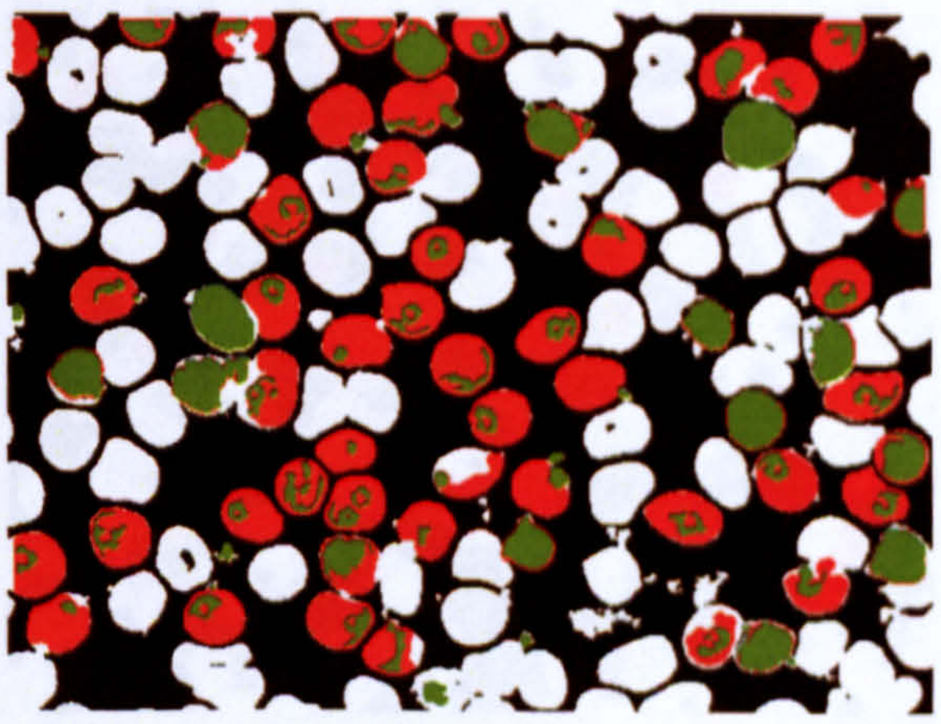
The stained pixel detection task is approached as a two class colour pixel classification problem. A Bayesian colour pixel classifier is implemented using the non-parametric density estimation method using histograms. A training set was used to construct the histograms for the stained and non-stained classes. The decision was made according



(a)

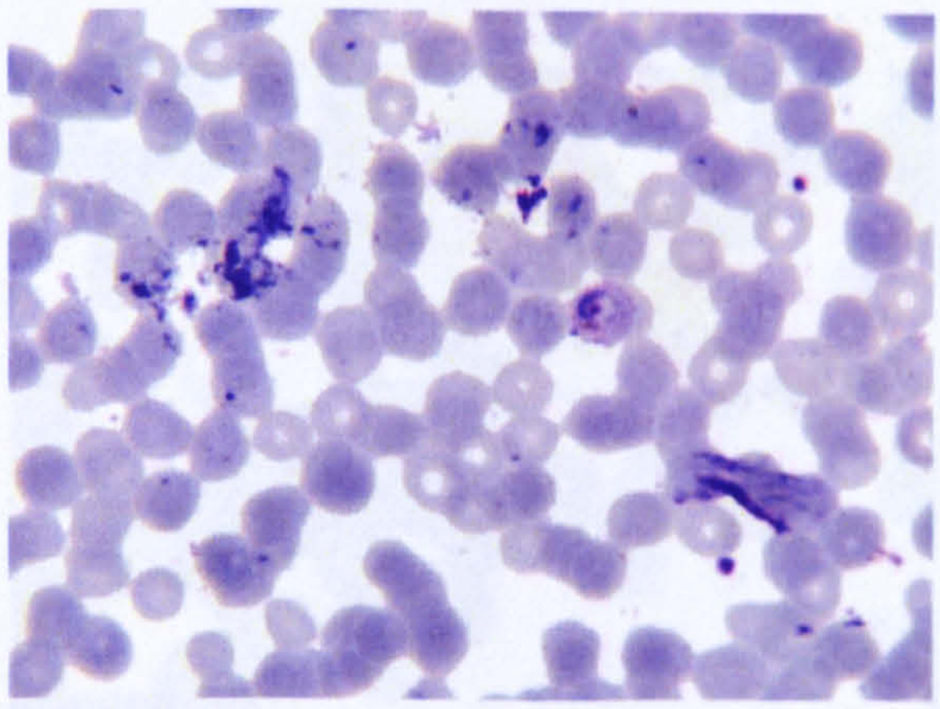


(b)

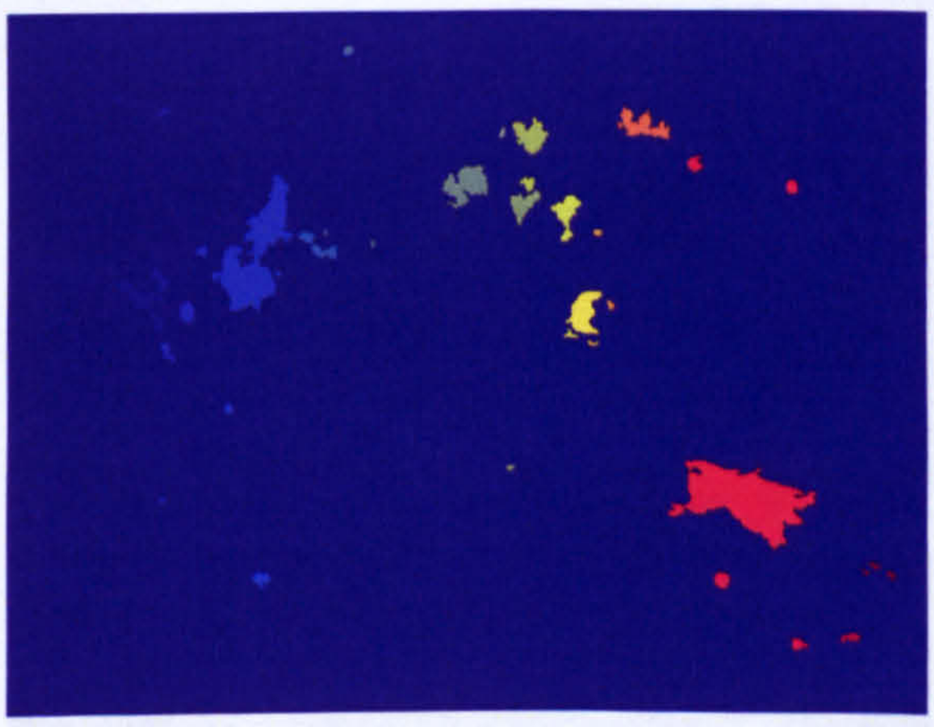


(c)

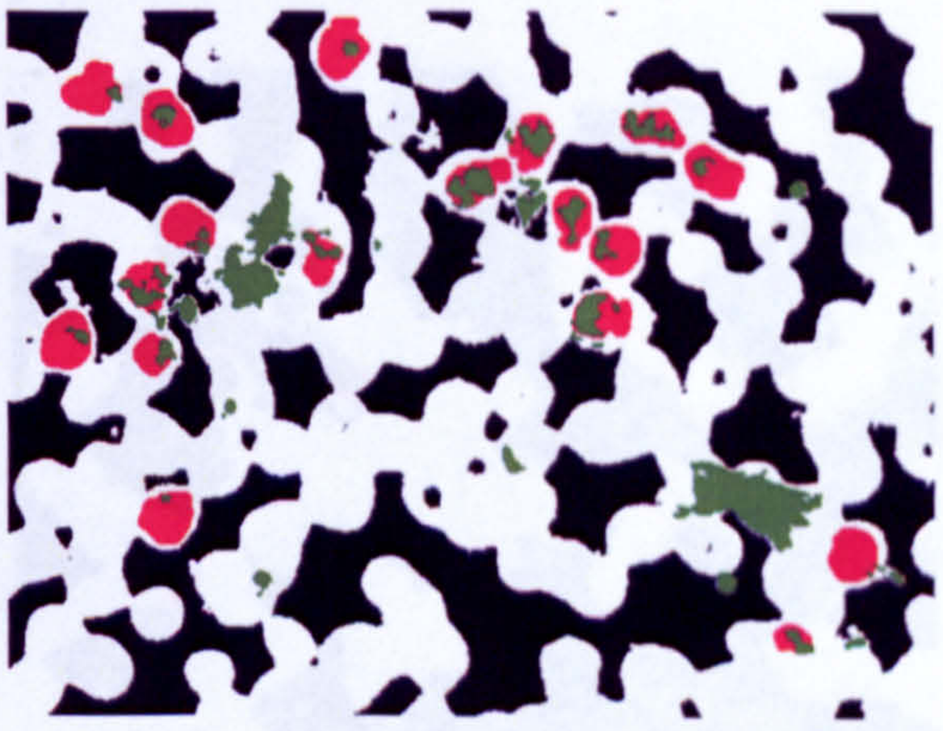
Figure 6.13: Stained object extraction example 1: (a) input image (normalised) from an *in vitro* image containing many *P. Falciparum* parasites, (b) labelled stained pixels, (c) extracted stained objects, red regions (obtained by α) denote the stained objects, green regions (obtained by β) denote enhanced stained pixels and initially found stained pixel groups.



(a)

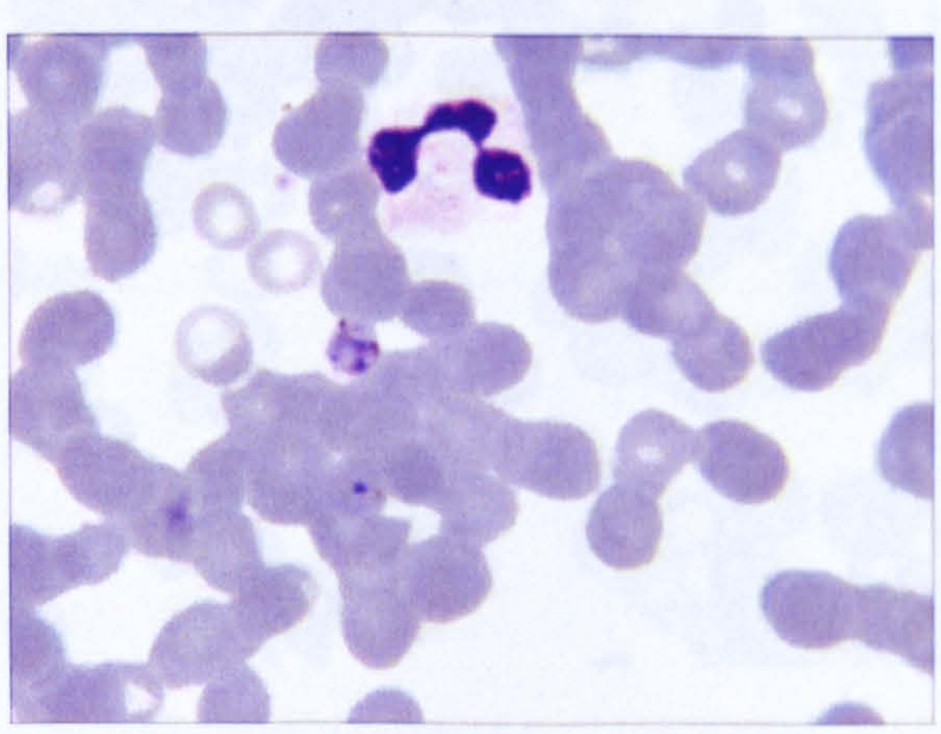


(b)

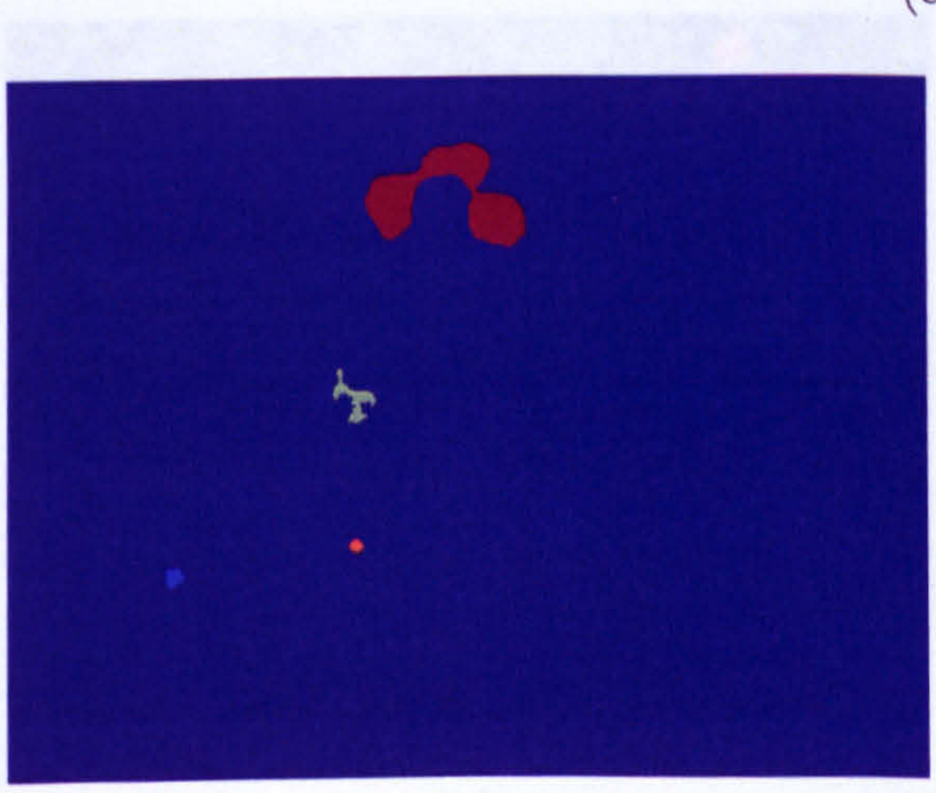


(c)

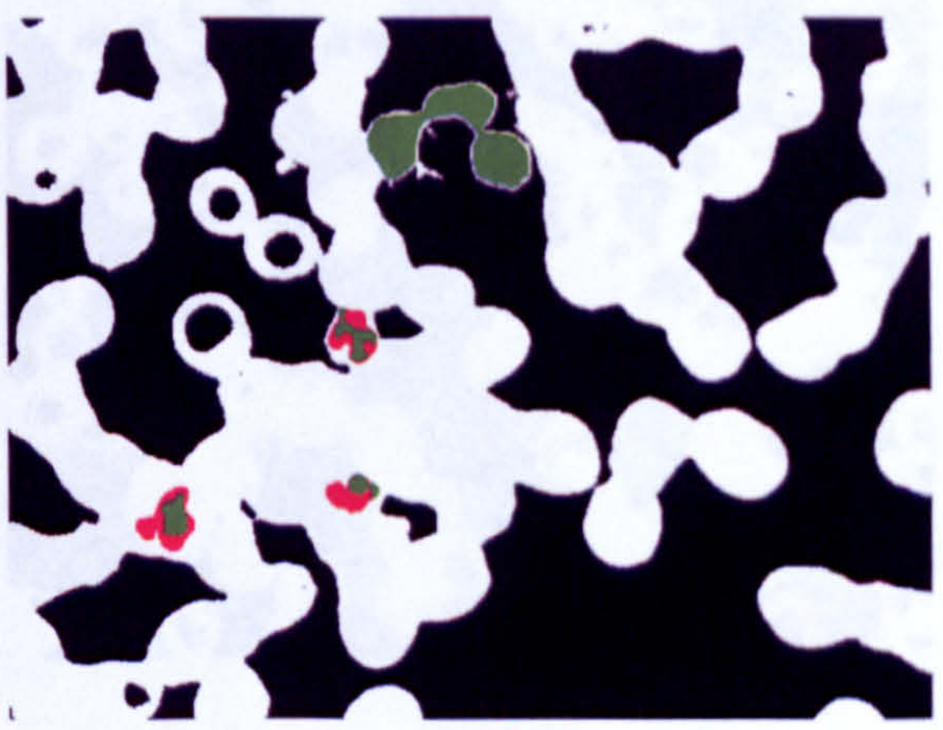
Figure 6.14: Stained object extraction example 2: (a) input image (normalised) from an dense field of peripheral blood slide containing one *P. Vivax* parasite (centre), (b) labelled stained pixels, (c) extracted stained objects, red regions (obtained by α) denote the stained objects, green regions (obtained by β) denote enhanced stained pixels and initially found stained pixel groups.



(a)

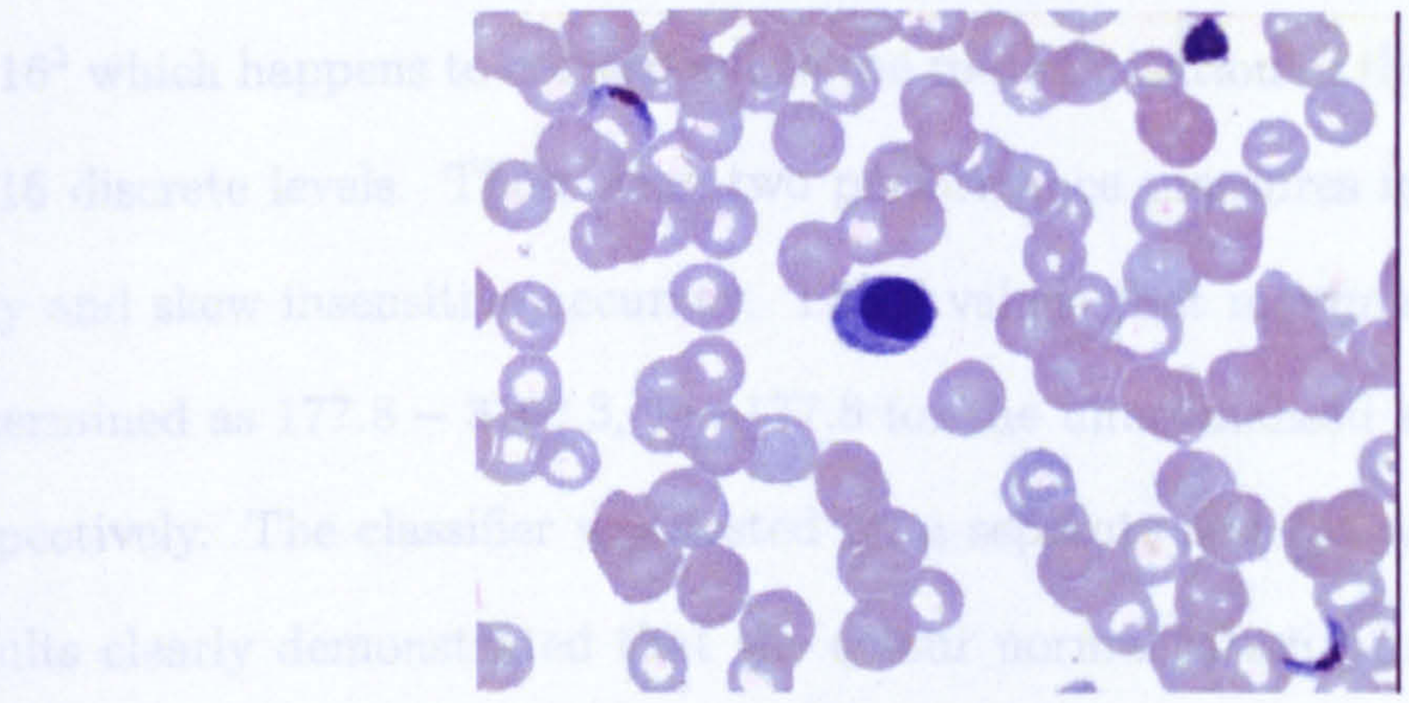


(b)

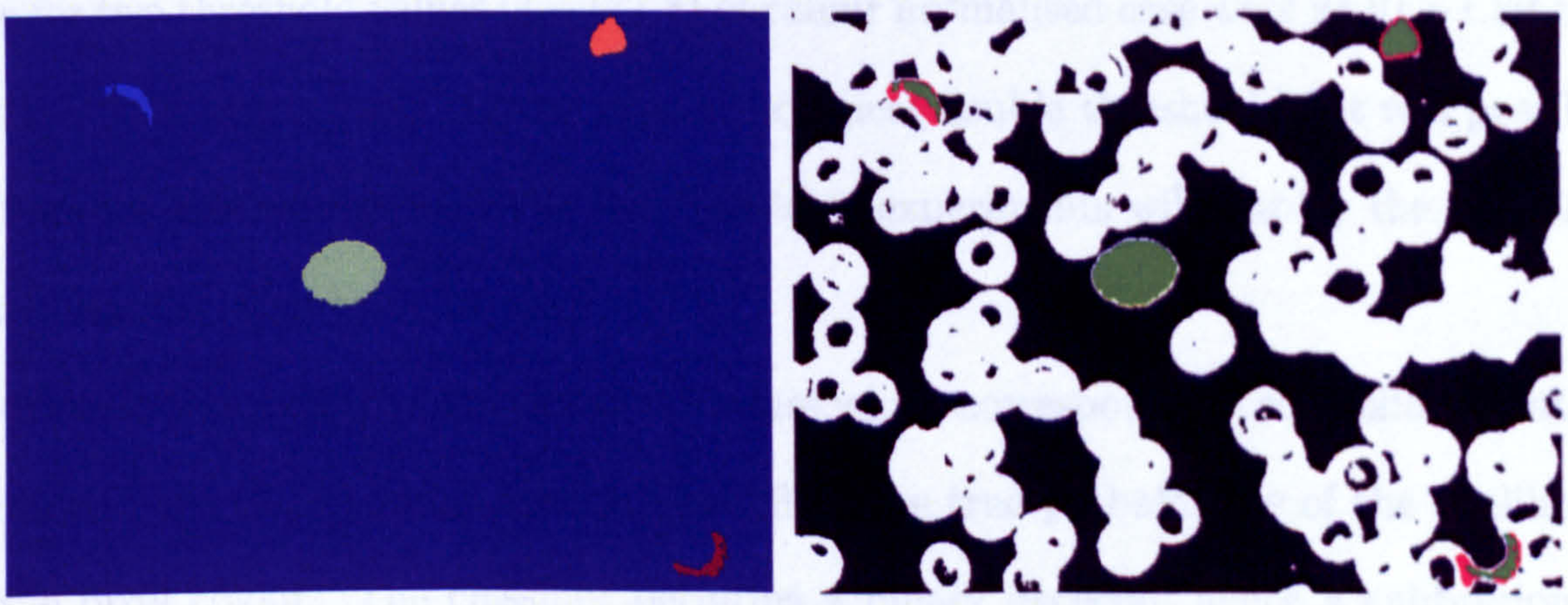


(c)

Figure 6.15: Stained object extraction example 3: (a) input image (normalised) containing two *P. Malariae* parasites (center two), (b) labelled stained pixels, (c) extracted stained objects, red regions (obtained by α) denote the stained objects, green regions (obtained by β) denote enhanced stained pixels and initially found stained pixel groups.



(a)



(b) (c)

Figure 6.16: Stained object extraction example 4: (a) input image (normalised) containing two *P. Falciparum* gametocytes (up left, bottom right corners), (b) labelled stained pixels, (c) extracted stained objects, red regions (obtained by α) denote the stained objects, green regions (obtained by β) denote enhanced stained pixels and initially found stained pixel groups.

to the likelihood ratio of the pixel colour ($[r, g, b]$). The likelihood threshold was determined by analysing the Receiver Operating Characteristics (ROC) curve of the optimal histogram size. The optimal histogram size was determined by comparing ROC curves of different histogram sizes $(256, 128, 64, 32, 16, 8)$. The optimal size was determined to be 16^3 which happens to correspond to the re-quantisation of the input colour channels to 16 discrete levels. There were two performance measures used: the average accuracy and skew insensitive accuracy. The θ values that maximise these measures were determined as $177.8 - 3162.3$, $1 - 177.8$ for the unnormalised and normalised images, respectively. The classifier was tested on a separate test set using these values. The results clearly demonstrated that the colour normalisation process improves the performance of the Bayesian classifier significantly. The true detection and false detection rate for two threshold values ($1 - 177.8$) of colour normalised case were $94.91 - 1.80$ and $83.12 - 0.08$, respectively. By using morphological double thresholding it was possible to improve the results ($94.2 - 0.74$). The later experiments will employ the Bayesian pixel classifier with these settings.

It must be noted that the threshold values which correspond to the likelihood ratios are not in the $([0, 1])$ range and do not reflect the true probabilities of the likelihood of the pixel colour. The classifier performs a binary decision; hence a calibration is not necessary. However, it is possible to calibrate the likelihood threshold to reflect the true likelihood probability. This can be performed by constructing a histogram of equal sized bins using all the sample likelihood values; then mapping them to a discrete sampled interval $([0, 1])$ [133, 134].

Additionally, the results were compared to the results obtained by a thresholding method that was proposed in [35]. It is shown that the new method outperformed the threshold method by providing a significantly higher true detection rate: $94.2 - 64.21$ but a slightly higher false detection rate: $0.74 - 0.25$. However, as discussed the higher false detection rate was also related to ground truth information which was used in the tests that was obtained by manual thresholding of the images (see Section 3.2.2).

The Bayesian pixel classification process has the advantage of using the supervised learning using the samples in the training and validation sets. The disadvantage is that the decision is based on a single pixel colour rather than a joint relation among all the pixels in the image, which is possible by histogram-based thresholding operations. It may be possible to develop a classifier utilising the input image characteristics. For example, two or more statistics from the image histogram could be used together with the RGB colour vector as features to produce the decision. However, histogram-based density estimation for a Bayesian classifier may not be proper for higher dimensional features if the samples are limited. Other classifiers such as Fisher linear discriminant or a neural network may be easier to train. This is identified as future work.

The stained object extraction algorithm is developed as an alternative to the top-down segmentation strategy. It does not depend on a top level assumption that all the objects in the images are RBCs and are segmentable based on the average cell area value. It uses the detected stained pixels as the starting point and tends to enlarge them by using area top-hat and reconstruction operations on the likelihood image. The algorithm extracts a region even when there are no resolvable object boundaries. In some cases this causes an arbitrary connected region to be extracted, however, limited by the area top-hat parameters. The values for the parameters were chosen heuristically. Thus, it is hard to suggest optimality.

One problem arises from the difficulty of assessing whether a stained pixel group has to be enlarged or not. A more comprehensive analysis is necessary to define the stained objects. Other image features such as edges, distance transform, local colour histograms can be utilised. However, if stained pixel detection can be improved, it may be possible to bypass the morphological operations which rely on the average cell area estimate. Therefore, the method would also be applicable to the thick film images and also other blood parasites. This is identified as future work.

The method presented here has been presented by the author in [36,37] within the parasite detection framework.

Chapter 7

Diagnosis - Methodology

An automated malaria diagnosis system requires two essential functions: parasite detection and species recognition. The parasite detection function analyses the object to determine if it is a parasite. The species recognition function analyses a detected parasite to determine its species category. In addition, to these, a third function, life-stage recognition, analyses a detected parasite to determine its life-stage, which can be used to assess the parasite development in blood. It is also important for laboratory researches (i.e. non-diagnosis applications) where response to a treatment must be monitored. This work approaches these problems from a statistical pattern classification perspective.

All the previous chapters discussed auxiliary functions to transform, seek and extract the objects for this analysis. Thus, they may be regarded as the tasks required for obtaining the object. It should be clear from the previous discussions that the above-mentioned tasks (i.e. parasite/non-parasite, species, and life-stage classifications) are not simple as the presence of artefacts in the images hinders the use of heuristic approaches.

This chapter explains the methodological basis of the parasite detection, species recognition and life-stage recognition experiments which will be introduced in the next two chapters (Chapters 8 and 9).

The following sections are organised as follows. Section 7.1 discusses the parasite and non-parasite object morphologies and identifies the general needs of the proposed classification schemes. Section 7.2 introduces the selected features, their implementation details, and some related issues, which are investigated in the classification experiments. Section 7.3 explains the feature concatenations and sequential forward floating feature selection algorithm of Pudil *et al.* [135]. Section 7.4 discusses some general classification schemes and explains the motivations for the selected classification algorithm.

Sections 7.5, 7.6, and 7.7 explain the performance measures, evaluation methods used in the experiments and the issue of the sample independence, respectively.

Readers who are familiar with these concepts may prefer to skip this chapter and return for reference as necessary.

7.1 Overview

In previous chapters, many thin blood film images are presented and several issues about the parasites and non-parasites are discussed. Avoiding a repeat as much as possible, it is still appropriate to define the terms parasite and non-parasite before beginning this discussion. The human *Plasmodium* species are referred to as "parasites". Since the objects that have reached this level of processing are ideally only the stained objects, then by the term "non-parasite" we refer to all the stained objects that are not *Plasmodium*.

Figure 7.1 demonstrates the sub-classes of the problem, which are subject to species and life-stage recognition tasks. It is possible to examine White Blood Cells (WBCs) with the sub-categories: Neutrophil, Eosinophil, Basophil, Monocyte, and Lymphocyte. Thus, the capability of the studied system can be extended. However, this is not an objective of this study.

The objects belong to the parasite class do not have a single morphology (i.e. appearance): There are four different species (*P. Falciparum*, *P. Vivax*, *P. Ovale*, *P. Malariae* and four different life-stages of interest for each species (*ring*, *trophozoite*, *schizont*, *gametocyte*). For Computerised Diagnosis of Malaria (CDM) applications this categorisation is sufficient; however for some other purposes the life-stages may be differently categorised (e.g. *merozoite*, *immature trophozoite/schizont*, *mature trophozoite/schizont*, *female/male gametocytes*). Figures 1.1, 1.2, 1.3, and 1.4 illustrate the different *Plasmodium* species in different life-stages. In the previous discussions, real parasite examples in thin blood film images were shown (e.g. Figures 6.1-6.4).

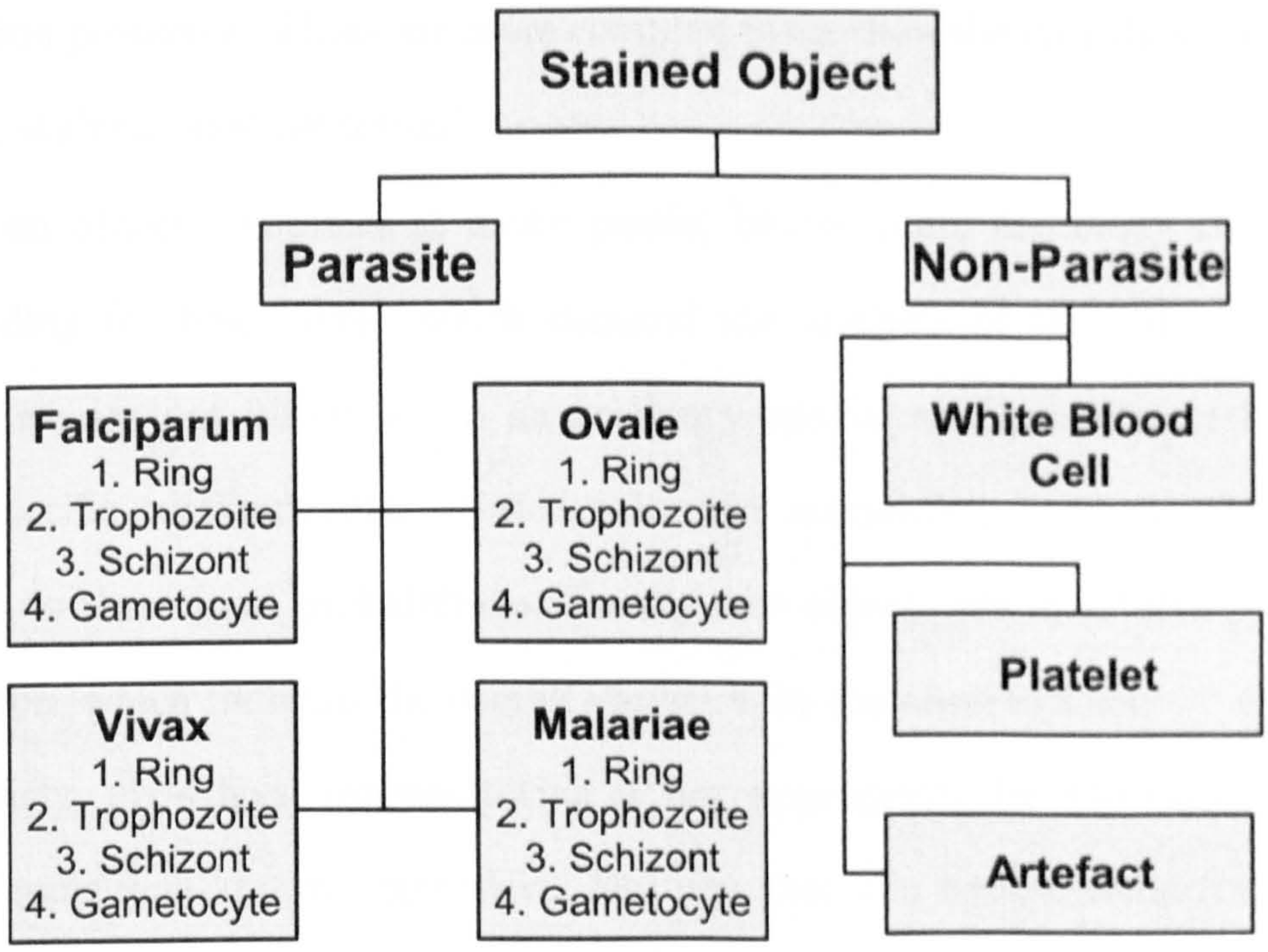


Figure 7.1: Hierarchy of classes in stained object classification

On the other hand, the non-parasite class has both categorised and not categorised components. The regular components such as WBCs and platelets are categorised and have defined morphologies; however, artefacts can not be easily categorised. An artefact can be due to Red Blood Cell (RBC) anomalies (e.g. anaemia), bacteria, spores, vegetable cells, fungus, dust particles, stain crystals, or defects in the slide glass. Therefore, the non-parasites are from a class, which has much broader variations in morphology.

For the parasite detection discussions, all the *Plasmodium* species and different life-stages are regarded as a single class: "parasite", whereas the other stained objects will be regarded as the second class; "non-parasite". On the other hand, for the species and life-stage recognition discussions, the parasite class is examined in the sub-levels as shown in Figure 7.1.

For parasite detection, we have a two-class pattern classification similar to the stained/non-stained colour pixel classification problem that was investigated in the previous chapter. For the species and life-stage recognition tasks, we have multi-class

classification problems. These are more complex tasks than the classification of a single pixel (e.g. stained pixel detection).

First, an object comprises of many pixels; hence, there are many colour vectors corresponding to these pixels, which demand the analysis of the joint probabilities. Secondly, we are not interested in an arbitrary spatial configuration of these pixels but the specific configurations (related to object morphology), which adds a spatial dimension to these joint probabilities. Finally, the objects are in arbitrary size, scale, and rotation, which increase the overall variation in the observations.

Obviously, pixel level representation is not appropriate for the tasks. However, there are some well-known higher-level features that can handle some (or all) of the mentioned variations. Hence, they can reduce the dimensionality of the problem and simplify the classification process. In this study, some very well known features such as histogram, moment descriptors, and colour auto-correlogram are investigated. In addition to these, the suitability of the less popular feature, area granulometry, is investigated. In the Section 7.2 the features are explained in detail.

Therefore feature vector dimensions will be higher than the simple RGB pixel colour feature vector that was used for the pixel classification task (Section 6.2); it is impractical to estimate the class conditional densities using the histogram approach and for the decision to rely on the likelihood as was the case for stained/non-stained pixel classification. However, it is possible to follow other non-parametric approaches (e.g. KNN, Parzen windows) to estimate the densities [33]. Alternatively, in these classification algorithms it is possible to estimate posterior probabilities or make the decision (i.e. class label) directly. In this study, the KNN scheme is used as a classifier. It directly produces the decision bypassing the density estimations. However, as discussed later, it performs implicit density estimation and posterior probabilities can be obtained. To provide a comparison some other classification schemes are also examined. The classifier architectures are discussed in Section 7.4.

Evaluating and assessing the generalisation capability of a classifier in a practical

application is not a straightforward task. The performance measures and the evaluation methods are discussed in Sections 7.5 and 7.6, respectively.

7.2 Features

In the literature Section 2.4 many studies which utilised different features and classifier architectures were reviewed for a very similar problem -WBC classification-. It was explained that various colour- and shape-based features and most commonly their different combinations were utilised as descriptive features for the WBC classification. In WBC classification, the problem is to determine the type of WBC (e.g. Neutrophil, Eosinophil, Basophil (polymorphonuclear) or Monocyte, Lymphocyte (mononuclear)). Hence, it is more similar to the task of the "species recognition" for malaria diagnosis. Compared to the parasites WBCs are more composite objects: in the case of most types, they can be segmented into sub-components such as cytoplasm and nucleus. Hence, ratios of some simple measurements on these two components (e.g. area, perimeter, or compactness) can be effective features for distinguishing the WBC categories. However, segmenting a parasite into several components is not possible, except for some mature stages.

The parasite is a non-rigid object that can have a large variation in its observed morphology (Figure 1.1-1.4). The colour distribution information is valuable but may not be adequate to distinguish the parasites from other stained objects. Raw images can not be used directly as features due to high variations in morphology, which are coupled with arbitrary rotations and scales. A feature to be used in this classification must provide translation, rotation, and scale-invariance and must be capable of capturing the morphological characteristics. In this study, although some of them are scale-variant, five different candidate features were chosen and implemented for this task: colour histogram, colour auto-correlogram, area granulometry, Hu moments, and a relative shape measurements feature vector. The following sections describe these features.

7.2.1 Colour histogram (H)

The histogram is a widely used descriptor, which is simple to compute and provides information about the colour distribution.

Suppose the input image I is quantised to have N distinct colours. Let c_i denote a single colour value ($c_i \in \{c_1, c_2, \ldots, c_N\}$). H is the count of the occurrences of the colour c_i in image I:

$$H(c_i) \triangleq \|I_{c_i}\| \qquad c_i \in C = \{c_1, c_2, \ldots, c_N\} \tag{7.1}$$

H can be normalised by the total number of pixels n to reflect the probability that a randomly chosen pixel will have the colour c_i: $h(c_i) = H(c_i)/n$ where h is the probability density function. Thus, H transforms to a scale-invariant feature (h) that represents the colour distribution of the object. However, in experiments the performance of both plain and normalised versions will be compared to provide a discussion. The length of the histogram feature vector is equal to N.

7.2.2 Colour Auto-Correlogram (C)

The colour auto-correlogram can be seen as an extended histogram: it carries spatial information in addition to colour distribution. It was proposed in [136] for image indexing and sub-region localisation and was found to be superior to the colour histogram feature on fixed size images. A slightly different implementation (i.e. the colour co-occurrence histogram) which can be described as the un-normalised correlogram was used in [137] for object recognition.

Correlogram is a joint three dimensional histogram, which considers pairs of colours together with their spatial locations on the image. Suppose the input image I is quantised to have N distinct colours. Let c_i and c_j denote colour values ($c_i, c_j \in \{c_1, c_2, \ldots, c_N\}$); and let D denote a predetermined distance set ($k_i > 0 \in D = \{k_1, k_2, \ldots, k_M\}$). Then Γ is the count of co-occurrences of pixels p_1, p_2 of colour c_i, c_j

which are k_i distance apart in I [136]:

$$\Gamma(c_i, c_j, k_i) \triangleq \| I(p_1) = c_i, |p_2 - p_1 = k| \, I(p_2) = c_j \| \tag{7.2}$$

This count can be normalised to reflect the probability that a randomly chosen pixel will have colour c_i and co-occur with a c_j colour pixel at distance k by dividing by $(8k \; H(c_i))$ which is the total number of pixels in distance k times $H(c_i)$ the number of pixels that have the colour c_i [136]:

$$C(c_i, c_j, k_i) \triangleq \frac{\Gamma(c_i, c_j, k_i)}{8k \; H(c_i)} \tag{7.3}$$

The $8k$ is due to the distance calculation with the city block measure (i.e. L_1 norm). For example, a centre pixel has 8 related pixels at distance $k = 1$, 16 related pixels at $k = 2$.

The colour auto-correlogram is a special constricted case of the above (general) correlogram definition where $c_i = c_j$ [136]. This constriction results in a two dimensional feature. Hence reduces the complexity of the calculation and feature length; and according to [136] without much loss in the representation capability. In this study, the colour auto-correlogram is implemented. For simplicity, henceforth it will be referred as the correlogram and will be denoted by C. The size of the correlogram feature is NxM.

The correlogram is a rotation- and translation-invariant feature; however it is not scale-invariant due to its fixed distance set (D) parameter. This issue is further discussed in the Section 7.2.6.

7.2.3 Area Granulometry (G)

In addition to the task of obtaining the size distribution of objects in an image, granulometry by opening with a set of fixed structuring elements has been used as a feature in many pattern discrimination applications [138–140]. However, area granulometry can be more appropriate as a discriminative feature than granulometry (by structuring elements) for the parasite detection task since it is not possible to generalise the

morphological structure of the stained objects by fixed geometric shapes (e.g. lines, squares, circles).

The area granulometry calculation for a grey level image was described as follows [45] (also see Section 4.2):

$$G_\Lambda(X) = \sum \gamma^a_{\lambda_i}(X) - \sum \gamma^a_{\lambda_{i-1}}(X) \qquad \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\} \tag{7.4}$$

The area granulometry of a stained object (in the grey level image X) represented by the binary mask S_b (see Section 6.4) can be calculated locally as follows:

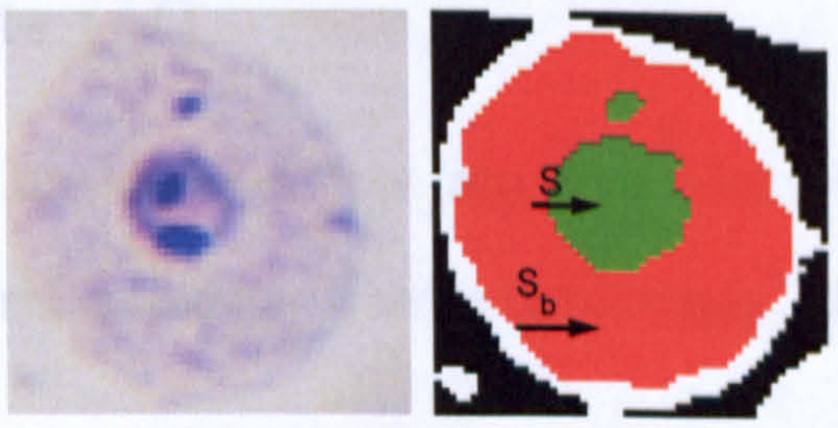
$$G_\Lambda(X. * S_b) = \sum \gamma^a_{\lambda_i}(X. * S_b) - \sum \gamma^a_{\lambda_{i-1}}(X. * S_b) \qquad \Lambda = \{\lambda_1, \lambda_2, ..., \lambda_n\} \tag{7.5}$$

where .* denotes pixel-by-pixel multiplication operation.

However, the granulometric distribution of the separate colour channels in an object may provide valuable information. If $G_\Lambda(X^c_{S_b})$ denotes the area granulometry of colour channel c of $X. * S_b$ where $c \in \{R, G, B\}$ then the area granulometry feature of $X. * S_b$ can be simply described as the set of the three calculations: $\mathbf{G} = \left\{G_\Lambda(X^R_{S_b}), G_\Lambda(X^G_{S_b}), G_\Lambda(X^B_{S_b})\right\}$.

The area granulometry feature ($\mathbf{G}$) is rotation and translation invariant; however due to calculation based on the intensity (i.e. differential total intensity), it is sensitive to the illumination unless it is normalised. Hence, a normalisation is performed by dividing the elements by the total intensity of the input X^c. In the case of the concatenated area granulometry of three separate channels, the granulometry from each channel is normalised separately with the sum of the intensities in the corresponding channel.

However, as in the case of the correlogram feature, the area granulometry is scale-sensitive because scale plays a part in its calculation. This issue is further discussed in the Section 7.2.6.



(a) (b)

Figure 7.2: An example stained pixel group within a stained object: (a) An example image (P. Vivax ring), (b) stained pixel group (S, green region(s)) and stained object (S_b, red region). In this case, the stained pixel group mask corresponds to the parasite and the stained object corresponds to the infected red blood cell.

7.2.4 Relative shape measurements (R)

The relative shape measurements vector, is formed from simple measurements to represent the stained pixel group shape relative to the stained object shape. Let S denote the binary mask for a stained pixel group and S_b the stained object (see Chapter 6.4) in which S is contained, $a(S)$, $p(S)$ are the area and the perimeter of S ($a(S_b)$, $p(S_b)$ for S_b), $m(S)$ moment of inertia of S, A_μ average object area in I (peak of the area granulometry of the whole image). An example stained pixel group within a stained object is shown in Figure 7.2. Note that both S and S_b are binary. Section 6.4 explains in detail how stained pixel groups and stained objects can be extracted from an input image. R is a six element vector containing the ratios below:

$$\mathbf{R} \triangleq \left\{ \frac{a(S)}{a(S_b)}, \frac{a(S)}{A_\mu}, \frac{a(S_b)}{A_\mu}, \frac{m(S)}{{A_\mu}^2}, \frac{4\,\pi\,a(S)}{(p(S))^2}, \frac{4\,\pi\,a(S_b)}{(p(S_b))^2} \right\} \tag{7.6}$$

R is a set of translation, rotation and scale-invariant features: $R(1)$ is the ratio of S and S_b areas; $R(2)$ and $R(3)$ are the relative ratios of area to the average cell area (that is measured by the area granulometry on the whole image). $R(4)$ is the moment of inertia of S that is normalised by the square of the average cell area. $R(5)$ and $R(6)$ are compactness measures of S and S_b respectively.

7.2.5 Hu moments (M)

Hu's moment invariants are derived from algebraic combinations of the first three orders of normalised central moments. The pq order central moment of a digital image is defined as

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q I(x, y) \tag{7.7}$$

where $\bar{x}$ and $\bar{y}$ are centres (mean values). The normalized central moments are derived from the central moments:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \quad \gamma = \frac{p+q}{2} + 1, \quad p + q = 2, 3, \ldots \tag{7.8}$$

Using nonlinear combinations of the second and third order normalised central moments, Hu [141] derived seven moments, which are translation, rotation, and scale-invariant.

Hence, M is a seven element feature vector, which includes these seven invariants. A complete description of the moments can be found in Hu's publication [141]. The moments of a stained object are calculated only for the stained pixel group binary mask S (see Figure 7.2).

7.2.6 Scale-Invariance

The distance values ($k_i \in D$) of the correlogram and the area threshold parameter ($\lambda \in \Lambda$) of the area granulometry are described in terms of pixel count and are not calibrated to reflect real physical measurements. For example, if we upsample an image by the factor of 2, the area granulometry using a chosen area threshold set (e.g. $\Lambda_1 = \{10, 20, \ldots, 100\}$ pixels) is not the same as for the initial image. However, a scaled set by 4 (e.g. $\Lambda_2 = \{40, 80, \ldots, 400\}$) will produce exactly the same curve (neglecting the effects of an interpolation).

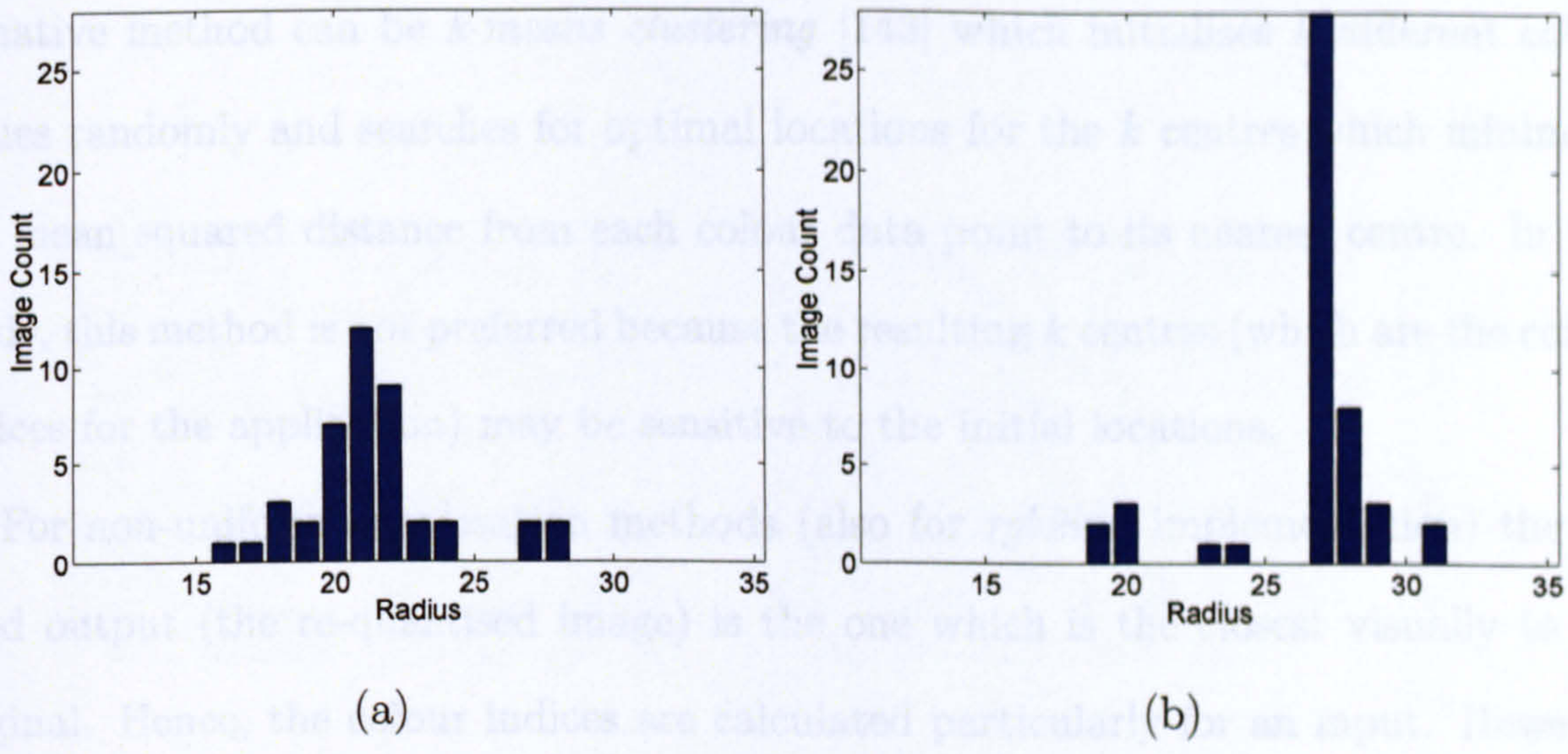
One guaranteed way of ensuring scale-invariance is to calibrate the pixel resolution according to the imaging system. Then a fixed set of threshold values for the area

granulometry or a fixed distance set for the correlogram which reflect physical measures can be safely used (e.g. $\Lambda_1 = \{20, 40, \ldots, 100\}\ \mu m^2$). However, this has to be performed in the early image acquisition stage. See the discussion in the Section 2.2.

An alternative way, which may reduce the effects of arbitrary scale, is to resample the images according to the area of a known blood component. An obvious candidate would be the area of RBCs: the average cell area measurement calculated by the area granulometry of the whole image can be used to estimate the relative resolution with respect to a reference image or a predefined value. However, in this case the RBC size variation in normal blood and some conditions, which cause abnormal RBC sizes would be neglected. Additionally, it must be considered that the thickness of the thin film through a slide varies which results in varying focus depths, which may slightly change the calculated average cell area. Therefore, the average cell area variation must be taken into account. Figure 7.3 shows average area distributions within two separate slides that are included in the University of Westminster (UoW) slide set. The images of second slide, which have the distribution shown in Figure 7.3(b) were captured under slightly higher magnification. Hence, the mean average radius (26) is slightly higher than the first one (21). This is a simple problem and can be corrected by scaling all the images from each input slide with respective to a constant radius. However, the estimated average radius variations within slides are significant. If every image is desired to be normalised (i.e. scaled) independently, the radius estimate should be assumed highly accurate. This issue is investigated in the experiments by comparing the independently scaled (according to the average RBC area) and un-scaled images.

7.2.7 Number of Colours

The colour based features, histogram and correlogram, use quantised fixed colour sets; the colours in the set are indexed (e.g. 1st colour, 2nd colour,..., Nth colour). Normally the images used in this study are quantised to have 256 distinct colours in each channel.



(a) (b)

Figure 7.3: The variation of area granulometry calculated average object (i.e. cell) area in two different slides: (a) and (b) show the area granulometry calculated average cell area distributions for two separate slides from the UoW thin blood film slide set. The x axis denotes the radius estimated by the average cell area (i.e. $Radius = \sqrt{Area}$), the y axis denotes the number of images having the corresponding radius estimate.

If all colours are used a 256^3 sized histogram or a 256^6xD sized correlogram would be produced. This is impractical for two reasons. Firstly, the feature space will suffer from the *curse of dimensionality* [131]: the limited number of samples will be insufficient to represent class distributions in such high dimensional spaces. Secondly, extraction, storing and comparing will consume a large amount of computational resources.

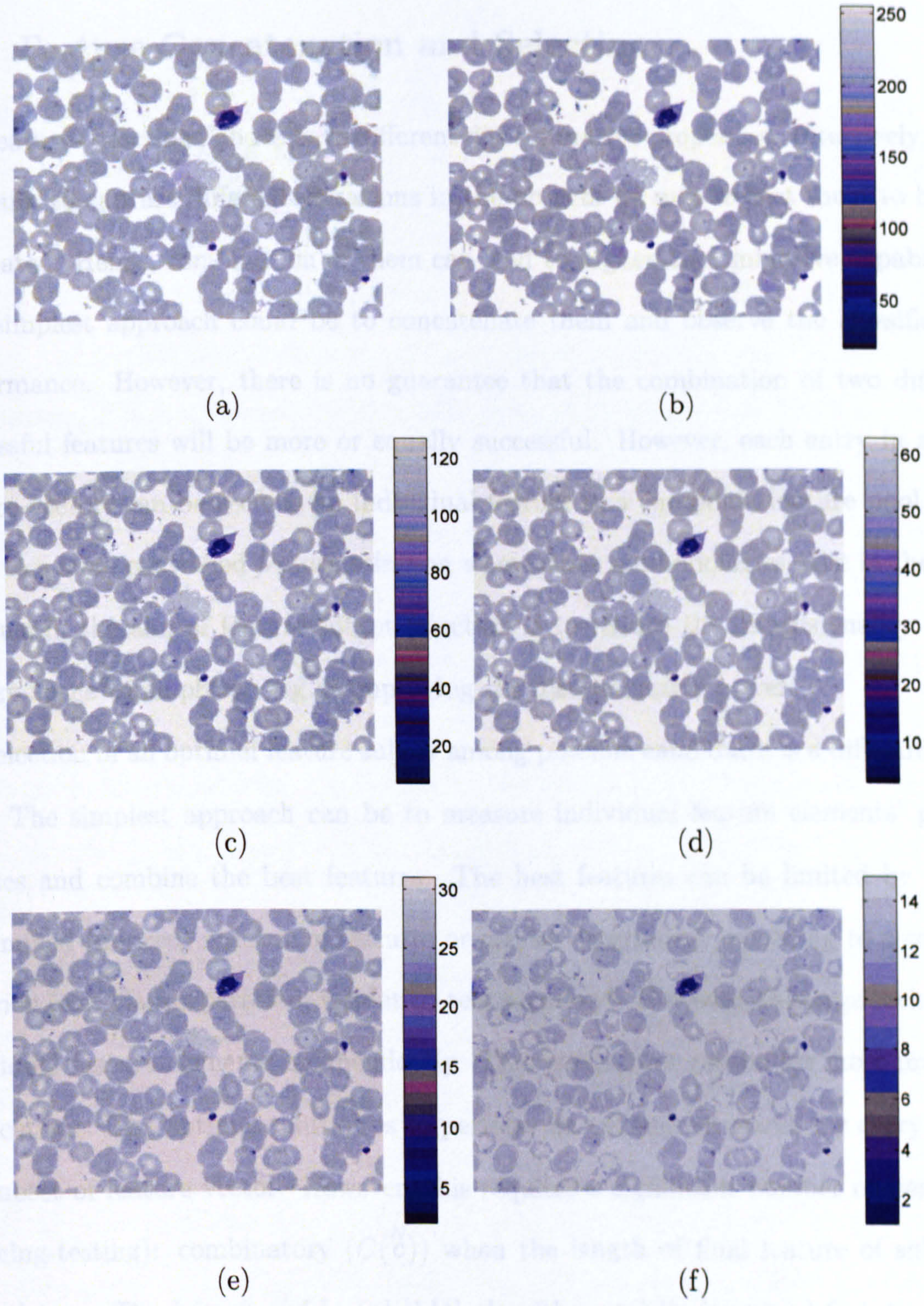
The same problem was faced in the pixel classification experiments; the solution was to re-quantise the images in order to reduce the dimensionality of the histogram size. The separate colour channels were independently re-quantised and new colour values were produced by uniformly sampling the original 256 distinct levels. However, this may be still insufficient to represent efficiently the colour distribution of a small sized object (e.g. 50x50 pixels). Therefore, a more efficient quantisation scheme is required. We have used the Matlab function *rgb2ind* to perform a non-uniform colour quantisation. The function implements minimum variance quantisation algorithm of Wu [142] to effectively reduce the number of total colours of a given image. An al-

ternative method can be *k-means clustering* [143] which initialises k different colour values randomly and searches for optimal locations for the k centres which minimises the mean squared distance from each colour data point to its nearest centre. In this study, this method is not preferred because the resulting k centres (which are the colour indices for the application) may be sensitive to the initial locations.

For non-uniform quantisation methods (also for *rgb2ind* implementation) the desired output (the re-quantised image) is the one which is the closest visually to the original. Hence, the colour indices are calculated particularly for an input. However, since our task is a classification we have to ensure the same colour indices of training samples are used for new queries. The colour indices can not be calculated for every image/object separately. Instead they must be calculated only once for the training samples and must be applied to all new queries. However, this process can be highly sensitive to colour characteristics of the images; and it is not desirable that the same colour index in one image refers to a different colour than another. Hence, colour normalisation is essential and should be performed prior to calculation of colour indices and re-quantisation of an input query image.

An example image and its re-quantised versions (with minimum variance quantisation [142], *rgb2ind*) for different numbers of output colours are shown in Figure 7.4. The colour-bars beside the images indicate the colour indices. The colour indices used to map this image were calculated from a set of objects that are collected from 20 different images. It can be seen that although the number of colours is dramatically reduced the re-quantised images are noticeably close to the original; it is hard to tell the difference between 128 distinct colour version and the original (~16 million). An obvious question that immediately arises is how to set the value (number of colours) for an efficient classification. This will be investigated in the experiments section.

It must be noted that indexed colour images are not suitable for mathematical morphology operations because the indices, which correspond to actual three dimensional colours, are not ordered. Therefore, although the area granulometry feature is



(a) (b)

(c) (d)

(e) (f)

Figure 7.4: Image re-quantisation for the colour based features: (a) The original input image (16 million colours), re-quantised versions (b) 256, (c) 128, (d) 64, (e) 32, (f) 16 colours.

calculated on colour images the same re-quantisation process was not applied.

7.3 Feature Concatenation and Selection

The features described above have different discriminative properties. Intuitively, since they aim to capture different variations in the objects we may expect them to be uncorrelated. Hence, combination of them can lead to higher discriminative capabilities. The simplest approach could be to concatenate them and observe the classification performance. However, there is no guarantee that the combination of two different successful features will be more or equally successful. However, each entry in all the feature vectors can be seen as an individual feature in a candidate feature pool. This leads to a more controlled *feature selection* where some methodologies exist in the field. In general, the aim of feature subset selection is to reduce the dimensionality of the feature space while preserving or improving the discriminative power.

Selection of an optimal feature subset among possible candidates is a difficult problem. The simplest approach can be to measure individual feature elements' performances and combine the best features. The best features can be limited by a predetermined (desired) number of features or can be determined according to a selected criterion (e.g. class separation, validation test accuracy). However, this approach treats individual feature elements independently. This yields poor results for most practical applications. The optimal solution is to perform an exhaustive search for every possible subset of feature vector. However, this requires a significant number of iterations (training-testing): combinatory ($C\binom{N}{c}$) when the length of final feature of subset is limited to c. The *branch-and-bound* [144] algorithm and its improved faster versions (e.g. [145]) preserve the optimality while iterating for only a fraction of possible subsets; however, they are still not efficient for many practical applications. The known alternative solutions are suboptimal. For example, in sequential methods (i.e. sequential forward/backward selection, SFS, SBS) the initial subset is increased/decreased

at each iteration by adding/removing the best/least successful feature. SFS can not exclude a feature once it is added to the subset whereas SBS can not include a feature once it is excluded. For example, a feature, which was found to be most successful and included in the first iteration of SFS may reduce the quality of the subset, which is formed in the subsequent iterations. This is known as the *nesting problem*. A solution, which is called *plus-l-take-away-r* is to use SFS and SBS in a conditional way to backtrack inclusions and exclusions as proposed by Kittler in [146]. Later in [135], the method was further improved to provide adaptive selection of l and r values: Sequential Forward Floating Search (SFFS). It is empirically shown that SFFS is the most reliable suboptimal method so far [147].

One critical issue in feature selection is the choice of the criterion, which assesses the quality of features. The *wrapper* models assume the quality of the features are classifier dependent [148]; and use the classifier as a part of the feature quality assessment. This can be performed by measuring the classification error on a limited set of samples for selected feature subsets. Depending on the complexity of the classifier, this can be impractical in many applications. On the other hand, the *filter* models assume that the feature quality can be assessed independent of the classifier [149]. Intuitively, the optimal set for a linear classifier and non-linear classifier can be expected to be different. However, under certain conditions (e.g. normal distribution) some criterion functions (e.g. Bhattacharya) are shown to be related to probability of error [150,151].

In this study, the concatenations of features are used to investigate whether the classification performance can be improved. Moreover, a wrapper model with the SFFS algorithm is used to investigate if any performance gain can be achieved by reducing the number of features in the concatenated feature vectors by discarding less effective and redundant features.

7.4 Classifiers

There are various parametric and non-parametric classification techniques, which can be used in the classification of the selected features. The parametric techniques assume that data distribution is in a particular form (e.g. normal distribution) and search for parameters which fit the data to distribution optimally.

On the other hand, there are various non-parametric (i.e. they do not assume a particular type of distribution) classification schemes. Most of the non-parametric methods do not assume a particular type of density but a particular form of a discriminant function. The discriminant function defines and operates the classification rule. For example, a discriminant function that performs the Bayes decision rule (see Section 6.2) for a two class problem (i.e. w_1, w_2) is in the following form:

$$f(x) = P(w_1|\mathbf{x}) - P(w_2|\mathbf{x})$$
$$\text{Decide} \quad w_1 \quad \text{if} \quad f(x) > 0 \quad \text{else decide} \quad w_2 \tag{7.9}$$

Thus, the feature $\mathbf{x}$ is assigned to the class with the maximum posterior probability ($P(w_i|\mathbf{x})\ i \in \{1, 2\}$). Most of the statistical pattern recognition techniques can be related to the Bayesian decision rule in terms of posterior probabilities. Therefore, using (7.8) they can be categorised as discriminant functions. However, we will regard the Bayes discriminant function as a special case since it has no explicit suggestion for the type of discriminant function.

A linear classifier (e.g. the Fisher Linear discriminant, perceptron [33]) can be used to search for an optimal decision line or hyperplane separating the two classes. A linear decision function (for a two class case) can be expressed as follows:

$$f(x) = \mathbf{w}^{\mathbf{T}}\mathbf{x} + b_0 \tag{7.10}$$

where $f(x) = 0$ defines the decision line in the feature ($\mathbf{x}$) space in the direction $\mathbf{w}$ with a bias (i.e. shift) value of b_0.

However, real world problems are unlikely to have linearly separable feature spaces. A non-linear classifier (e.g. multilayer neural networks) can be trained to estimate a

non-linear function, which can divide the feature space non-linearly. For example, multilayer neural networks with sufficient hidden layers and units are well known to estimate any complex decision function [32]. A decision function for a two layer fully-connected single output network can be in the following form:

$$f(x) = g(\overbrace{\sum_{j}^{M} w_j \, g(\underbrace{\sum_{i}^{N} w_{ji} x_i + b_{i0}}_{1}) + \, b_{j0}}^{2}) \tag{7.11}$$

where $g(.)$ denotes a non-linear activation function (e.g. sigmoid, hyperbolic tangent) which provides the non-linearity. The multilayer structure provides the expressive power by first (1) projecting the input feature vector ($\mathbf{x} = [x_1, x_2, ..., x_N]$) onto each of M directions described by (w_{ji}) and hidden unit biases b_{i0}; followed by the non-linear mapping $g(.)$ which is then projected onto another set of directions described by (w_j) and the output unit biases (b_{j0}); finally it is mapped non-linearly to the output by $g(.)$. The non-linear activation function ($g(.)$) can be chosen differently for the different layers and units.

Another approach is to apply a non-linear mapping to feature space which then can be divided by a linear discriminant decision function (e.g. kernel methods, non-linear support vector machine) [31, 33]. A generalised linear discriminant function is linear however not in the original feature space:

$$f(x) = \mathbf{w}^{\mathbf{T}} \phi + b_0 \tag{7.12}$$

where ϕ is a vector function which performs a non-linear mapping of $\mathbf{x}$ (e.g. $\phi_1 = x_1^2, \phi_2 = x_1 x_2, \phi_3 = x_2^2$).

A supervised learning or training method for a discriminant function varies depending on the type of function: however, the common principle is to search for parameters, which minimize the difference between the desired outputs and the function outputs for a set of input examples (i.e. the training set). After the training is complete, the examples are discarded. The discriminant function with the estimated parameters represents the classifier, which is a division of feature space for the decision.

7.4.1 K-Nearest Neighbour Classifier

In the literature, a popular classification method does not assume a particular density function form or a discriminant function type: *K-Nearest Neighbour rule* (KNN) [152]. Usually categorised by text books [31–33] as a non-parametric density estimation method. KNN bypasses the density estimation and directly produces the posterior probabilities and the decision.

A simpler form of KNN, the *nearest neighbour* classifier assigns a query vector to the class of the nearest one in the set of the previously seen examples (i.e. training set). Hence, it can be shown that the decision boundaries formed by the nearest neighbour classifier correspond to Voronoi partition of the feature space [33]. An obvious extension of the nearest neighbour rule is KNN which assigns a query vector to the class which has the majority (most frequently represented) among the K nearest neighbours in the set of the previously seen examples. The decision surface formed by a KNN classifier (when $K > 1$) is not fixed; it is formed locally around the query. Hence, it can be described as a *lazy learning* approach where the learning/reasoning is performed per instance. This is the opposite case of previously described discriminant function approaches where the learning is performed before seeing the query.

Let $\mathbf{x}$ show the query vector for a KNN classifier that is built with N labelled samples. Suppose a volume V is placed around $\mathbf{x}$ and K neighbours are captured among which k_i are from class w_i. Then it can be shown that an estimate for class-conditional probability density $p(x|w_i)$ and posterior class probability can be written as follows [33]:

$$p(\mathbf{x}|w_i) = \frac{k_i}{N\,V} \tag{7.13}$$

$$P(w_i|\mathbf{x}) = \frac{k_i}{K} \tag{7.14}$$

The training procedure of a KNN classifier is simply to store the training samples together with the corresponding class labels. If a query is presented the decision is made by the votes of K nearest neighbours. Intuitively, it can be expected that the

accuracy of the classifier will increase as K and the total number of training samples (N) increase. Theoretically, it can be shown that K-nearest neighbour rule error rate approaches to the theoretical minimum error (Bayes error) as both N and K approach to infinity and $K/N \rightarrow 0$ [33, 152].

The KNN decision rule has attracted many research studies due to its simplicity and its comparable performance to much sophisticated classifier architectures. The KNN classifier differs from the discriminant function based classifiers in some certain aspects. First, the KNN classifier is built (not trained) by training samples which is relatively a simple task. However, to produce a decision for each new query the distances for all of the training samples have to be calculated which could be impractical for large problems. There are various studies concerned with speeding-up KNN classification by either reselecting the data or reorganising them to improve efficiency. For example, by condensing and editing operations the samples, which contribute to misclassifications, can be removed from the training set to reduce the number and improve efficacy [33]. Alternatively, the computational load can be reduced by using search trees or partial distances [153–155].

7.4.1.1 Summary

In summary, there are several reasons that KNN is chosen as the classifier for the task:

- It is simple, intuitive and proven to be successful in many practical applications.
- It has a single parameter K which is in fact determined by data set itself; it does not require exhaustive parameter searches.
- It is a multi-class classifier. The interesting property is that the same classifier can be used for binary and multi-class classification. This property enables us to perform parasite detection and species recognition by revealing the target class sub-labels in the training sets.

- It is relatively simple to extend the classifier's capability to diagnose different diseases, which is a desirable property from the CDM point of view. For example, another parasite for a different disease can be included in the future by simply adding new samples instead of training the system once again. However, it should be provided with the same features as long as they are discriminative for the parasites (or elements) of the new disease.

KNN classification is very simple and it will most likely produce comparable results to more complex learning algorithms regardless of settings. However, there are some details reviewed below which can slightly affect the performance.

7.4.1.2 The K value

In Section 6.2 we have utilised histograms for non-parametric density estimation. The histogram bin size was corresponding to a smoothing parameter for the density functions. In a similar way, the value of K in KNN classifier is a smoothing parameter for density estimation. Therefore, larger values of K are desirable to obtain a reliable estimate. However, in the finite sample case, as $K \rightarrow N$ the conditional posterior probability of the classes will approach the probability of the class itself within a finite sample set (i.e. $P(w_i|\mathbf{x}) \rightarrow N_i/N$, where N_i is the number of samples of class i in total number of samples N). Hence the K value should be a small fraction of N. In practice, the value of K is usually determined empirically by cross-validation. In [156] Wettschereck *et al.* proposed that local, per-class or per-cluster K values can be used instead of a single global K value. According to the experiments performed, the conclusion was that local K values may improve the KNN on *specific* datasets.

7.4.1.3 Distance Measures

Since the decision for a query is made by votes of the K *nearest* sample points in the feature space, the most important aspect of the KNN decision is the definition

of distance. In the simplest case the Euclidean distance metric (L_2) can be used to calculate distances. Other simple choices for a simple measure can be the L_1 norm (L_1), a relative distance measure based on L_1 norm (RL_1) which is a variant of the Canberra metric [31,136].

$$L_2(\mathbf{x},\acute{\mathbf{x}}) \triangleq \sqrt{\sum_{i=1}^{n}(x_i - \acute{x}_i)^2} \tag{7.15}$$

$$L_1(\mathbf{x},\acute{\mathbf{x}}) \triangleq \sum_{i=1}^{n} |x_i - \acute{x}_i| \tag{7.16}$$

$$RL_1(\mathbf{x},\acute{\mathbf{x}}) \triangleq \sum_{i=1}^{n} \frac{|x_i - \acute{x}_i|}{\mu + |x_i| + |\acute{x}_i|} \tag{7.17}$$

The L_2 (Euclidean distance) is known to be less robust against outliers than the L_1 (city block distance) [33]. However, both L_1 and L_2 metrics equally weight all dimensions of the feature space. Therefore, they imply that the feature space is isotropic. This is not applicable to many practical problems [157]. For example, a feature vector, which is formed of the perimeter and area measurements, is not isotropic, because feature axis would be differently scaled. Suppose that the feature pairs are from different axes in the feature space, and the first one is scaled by a factor of 100 ($(10,20) * 100$). The result of L_1 will change (1000) drastically. On the other hand, the RL_1 result would be more stable since it performs a normalisation and reduces the effect non-isotropic feature spaces by compensating for relative scale differences. The constant μ value in the denominator prevents division by zero but prevents RL_1 from being completely symmetric with respect to origin. It should be set to a value, which is smaller than possible feature values. In RL_1 the contribution of the differences in lower feature values is relatively more significant. Suppose we calculate the distance using L_1 for two pairs of feature values: $(10,20)$ and $(1000,1010)$. Although, the difference seems much significant for the first pair, the L_1 result is 10 for both. On the other hand, the RL_1 result is higher for the first pair. A visual demonstration of the different metrics is provided in Figure 7.5.

The above distance measures assume feature vectors are aligned and they only con-

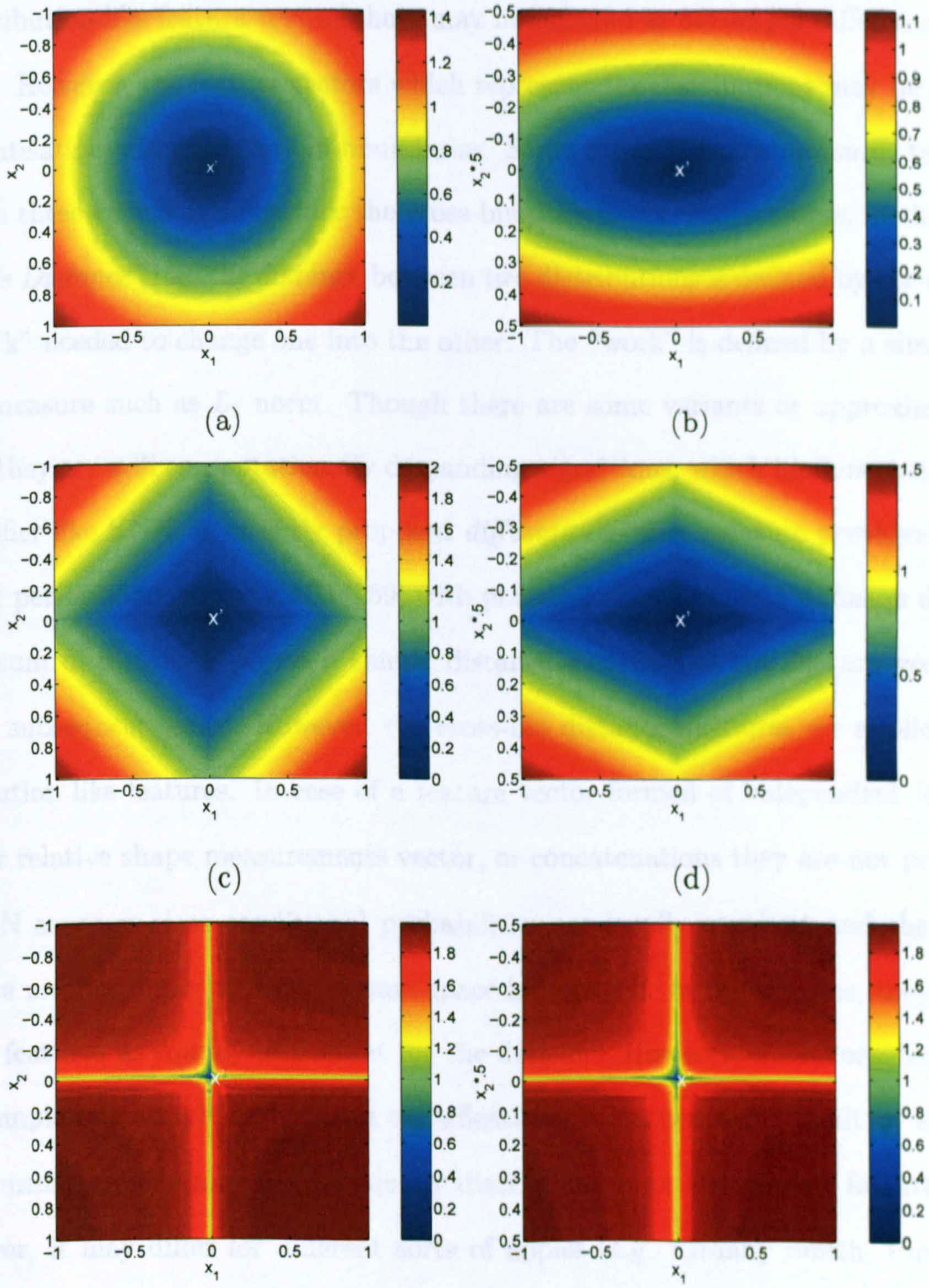


Figure 7.5: Distance metrics: (a) Euclidean distance (L_2 norm), (c) city block distance (L_1), (e) relative L_1 norm (RL_1) calculated on positive valued two dimensional space. (b), (d), (f) demonstrates the same calculations with the same order when x_2 axis is scaled by a factor of 0.5. The points of distance calculations are shown with $\acute{x}$.

sider differences between two corresponding feature elements. In the case of histograms or distribution-like feature vectors these may be referred as *bin-to-bin* difference calculations. However, the feature vectors which represent the distributions may be subject to quantisation effects in the bin boundaries. Some other distance measures take into account these effects and consider the cross-bin differences. For example, in the *Earth Mover's Distance* [158] the distance between two distributions is defined by the amount of "work" needed to change one into the other. The "work" is defined by a simple distance measure such as L_1 norm. Though there are some variants or approximates of EMD, they are still computationally demanding algorithms, which hinders their use in a classifier like KNN. A recently proposed *diffusion* distance measure provides a comparable performance to the EMD [159] with practical efficiency. The diffusion distance is the sum of distances (again a simple distance measure) of two feature vectors in several subsequent scales. However, the cross-bin distance measures are applicable to distribution like features. In case of a feature vector formed of independent features, like our relative shape measurements vector, or concatenations they are not practical.

KNN assumes class conditional probabilities are locally constant and the simple distance metrics imply that the feature space is isotropic. In other terms, they regard all the features as equally important for the decision. However, if we consider a simple example (e.g. apple and orange classification), a feature vector built by size and colour measurements may not be equally discriminative in all over the feature space; moreover, it may differ for different sorts of apples (e.g. Granny Smith, Pink Lady or Gala) and oranges (e.g. Mandarins, Oranges, Tangerines). A simple approach to the problem can measure the individual feature performances separately and use the results as feature (relevance) weights in the distance calculation. This process may reduce the overall error rate; however still it is global and has no solution for an inhomogeneous feature space. Alternatively, after K nearest neighbours are calculated, a local operation can be used to morph distance calculation locally around the query point: promote more relevant, and demote less relevant features. In [160] Hastie *et*

al. performed local linear discriminant analysis to estimate local feature relevancies, which are then used for weighting the contribution of elements in the local distance calculation. The same approach is followed in some other studies using support vector machines [157,161]; learning vector quantisation [162]; and *Chi-squared* distance [163]. The adaptive distance measures are shown to improve KNN classifier performance on some common data sets [163]; however, these are not implemented in this study and are left for future studies.

7.5 Performance Measures

In order to analyse classification performance, general pattern (or signal) detection task naming conventions can be followed as for the pixel classification task (Section 6.2): measure the true detection versus false detection percentages. However, in this case, given that users of this work are likely to be medical practitioners or researchers, it may be more appropriate to use the medical diagnosis naming conventions to help interpretation of the results.

Independent of the test result (classification) a test sample can be *positive* (infected) or *negative* (not infected), as determined by another (reference) reliable method. There can be four different outcomes after the classification: (tp) true positive (the classification result is positive for a positive sample); (tn) true negative (the classification result is negative for a negative sample); (fp) false positive (the classification result is positive for a negative sample); (fn) false negative (the classification result is negative for a positive sample). The number of occurrences of these conditions can be used to analyse the diagnosis performance: TP, TN, FP, FN for tp, tn, fp, fn, respectively. Calculated measures and interpretation of them for our experiments are as follows (see also [164]):

- **Sensitivity** is the proportion of the samples that are classified as positive among all the positive samples. For our task, it is the probability of a positive result

among infected RBCs.

$$SE = \frac{TP}{TP + FN} \tag{7.18}$$

which is usually called the true detection rate (in the pattern recognition community) for binary pattern detection tasks. The higher the sensitivity, the less likely that an infected RBC will be missed. Consequently, an infected person is more likely to be diagnosed as sick.

- **Specificity** is the proportion of the samples that are classified as negative among all the negative samples. It is the probability of a negative result for a negative object (regular blood component, e.g. RBC, WBC, platelet, and artefact).

$$SP = \frac{TN}{TN + FP} \tag{7.19}$$

The higher the specificity, the less likely that a healthy blood component will be classified as a parasite. Consequently, a healthy person is more likely to be diagnosed as healthy. Sensitivity and specificity values of a diagnosis test should be interpreted together. In theory, for an ideal diagnosis method the values are independent and both can be as high as 1.0. However, in practice a trade-off is usually observed. A useful diagnosis test should ensure at least one of the values is very high (e.g. > 0.95). However, it should be noted that assigning all the queries to one of the classes (consider the positive class) can simply achieve this ($SE = 1.0$) but the specificity in this case would be $SP = 0.0$ which is not desirable.

In general pattern detection tasks, the value $(1 - SP)$ is usually called the false detection rate, which can be used together with SE (true detection rate) to plot the receiver operating characteristics of the classifier (See Section 6.5).

- **Positive Prediction Value** is the proportion of the positive samples of all that are classified as positive.

$$PPV = \frac{TP}{TP + FP} \tag{7.20}$$

Positive prediction value indicates the reliability of a positive result. It answers a very intuitive question: what is probability of the patient being sick given the diagnosis is positive.

- **Negative Prediction Value** is the proportion of the negative samples of all that are classified as negative.

$$NPV = \frac{TN}{TN + FN} \tag{7.21}$$

Negative prediction value is the opposite of PPV. It answers the question: what is the probability of the patient being healthy given the diagnosis is negative.

Sensitivity and specificity values are not affected by the population of positive and negative samples among the test set (i.e. sample distribution). However, prediction values are sensitive to the sample distribution.

The average accuracy, which is the total number of misclassifications over total number of samples, can be expressed as follows:

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} \tag{7.22}$$

As seen in the pixel classification experiments (Section 6.5), average error or accuracy is not very informative when the examples among different classes are not equally distributed (i.e. skewed) in the sample set. The skew insensitive accuracy measure [49] can be expressed in terms of the sensitivity and specificity measures: $(Acc* = (SE + SP)/2)$. It can be used as an additional or alternative measure for unbalanced sets. In the experiments, both of the measures are utilised. The average accuracy is denoted by Acc; skew insensitive accuracy is denoted by $Acc^{\bullet}$.

7.6 Evaluation Methods

In practical pattern recognition, we generally would like to estimate the accuracy of the overall procedure by performing tests on a concrete set of samples. The choice

of the testing procedure and error measures can significantly affect the results (i.e. error estimation) [165]. However, the factors affecting the results are complex and unfortunately, the true error rate, which may show the accuracy of the estimation, is unknown. There are several commonly used testing procedures in pattern recognition.

In the resubstitution method, a set of samples are used in both training and testing. This method is significantly optimistic. However, it can be used to estimate a lower bound for error. In the hold-out method, the set of samples is divided into two (training and testing). In the m-fold method, the set is randomly divided into m parts and the $m-1$ parts are used for training while the rest is used for tests. This procedure is repeated m times and the results are averaged. In the case that m equals to N (number of samples) the method is referred as leave-one-out method. In this method, every sample is used once for testing while the remaining data is used for the training. This is a more statistically justified method for tuning and testing than the resubstitution and hold-out methods. However, it corresponds to performing the training process N times for a set of $N-1$ samples which can be impractical for the most classifiers with complex training processes.

7.7 Sample Independence

In general, the hold-out or leave-one-out evaluations ensure the independence of the training and testing samples. Hence, a success using the independent samples suggests a generalisation in real applications. However, the sample independence should be more clarified in order to discuss the capabilities of a potential CDM system. Specific to the CDM application, we can define different levels of sample independence:

- Different Samples: The basic and necessary level. The test results may indicate a generalisation capability.
- Different Images: Enables varying scales and colour characteristics to be tested. The test results may indicate a generalisation capability under different image

conditions.

- Different Slides: Enables varying slide preparation conditions such as different stain concentration exposures to be tested. The test results may indicate a generalisation capability under varying slide conditions. Different blood samples can also be seen under this degree of independence.

- Different Imaging Systems: Enables thoroughly independent samples to be tested. The test results may indicate an imaging system independent generalisation capability. Therefore, the results can be expected to be the same for the different places where the system is employed as a diagnostic tool.

Chapter 8

Diagnosis - Parasite Detection

The most important topic of Computerised Diagnosis of Malaria (CDM) from this thesis' point of view is the *detection* of parasites. It aims directly at the primary task for the diagnosis of malaria infection: determining whether a person's blood is infected by malaria parasites. However, if the result is positive a further classification is required to specify the malarial parasite.

This chapter investigates the suitability of various colour and shape based features and their concatenations for the parasite detection task. Additionally, a feature selection algorithm (Sequential Floating Forward Selection [135]) was employed to investigate whether any improvement can be achieved by reducing the feature vector size. The decision rule that is used for the classifications is based on the K-Nearest Neighbour (KNN) algorithm [33]. However, a comparison with two other classifier architectures (Fisher Linear Discriminant, Back Propagation Neural Network) is provided.

The methodological explanations of the experiments can be found in Chapter 7. The main contributions to the state of the art, introduced in this Chapter can be listed as follows:

- The detection problem is comprehensively investigated by highlighting the important practical issues for the realisation of the CDM system.

- A classification scheme is proposed for parasite/non-parasite classification (i.e. parasite detection). Therefore, a realistic malaria diagnosis tool for thin blood film examination is produced. The proposed scheme can differentiate between a parasite and regular (stained) blood component.

- The findings of the parasite detection study was published by the author in [36] and [37] with narrower contexts.

8.1 Introduction

In two earlier publications [36, 37], we have used 260 images of four thin blood film slides including P. Falciparum, P. Vivax, P. Ovale infections. The previously described features (i.e. histogram, Hu moments, relative shape measurements, and correlogram) except the granulometry feature and a KNN classifier were used to evaluate the parasite detection performance. The individual and concatenated feature performances were tested. The shape based features (Hu moments and shape measurements vector) showed a poor performance; however they slightly improved the performances of the colour based features (correlogram and histogram) when a concatenated feature vector was used. The best *hold-out* test performance was obtained using the combination of the correlogram and shape based features: ~74% sensitivity, ~98% specificity, ~88% positive prediction, and ~95% negative prediction values for the parasite detection (parasite/non-parasite classification). The performance of the concatenation of the histogram and shape based features was similar. Additionally, we have shown that the stained pixel detection or stained object extraction as proposed in [34, 35, 73] does not lead to successful parasite detection; parasite detection can also not be performed with heuristics based only on area measurements (thresholding based on the area of the objects).

The experiments described in the following sections used an extended data set of 630 images (including the former set) from 9 slides including *P. Malariae* (henceforth, Malariae) parasites. The ground-truth information was provided by manually labelling the stained objects in the images (outputs from the stained object extraction process, see Section 3.2.3). The number of objects in the parasite and non-parasite classes was 669 and 3431 respectively; where 336 and 1786 parasite and non-parasite objects were used for the training (Ta) and 333 and 1645 objects used for the testing (Tt), respectively. The images were separated to ensure the objects used in the training and testing were from different images. However, due to the limited number of slides,

different objects from the same thin blood film slide had to be separated into both the training and testing sets. See the sample independence discussion in Section 7.7. Finally, all of the images were pre-processed with the colour normalisation algorithm presented in Chapter 5 to provide colour constancy.

The experimental methodology is also extended to provide more insight into the problem and the proposed method. The feature performances are comprehensively experimented; dependences on the feature extraction parameters are investigated. The parameter selections were performed according to the skew insensitive average accuracy measure of the *leave-one-out* evaluations (see Section 7.6) on the training set. As explained in Section 7.5, the $Acc*$ measure provides a better (more balanced) accuracy estimate in unbalanced sample sets. However, as will be seen latter, the average accuracy measure (Acc) was also used when necessary.

Exhaustive parameter searches and intensive tuning were avoided since they could lead to over-optimised results for the dataset. The strategy for the experiments can be summarised as follows:

- The whole set was divided into two separate sets: the training (Ta) and testing Tt as described before.
- The leave-one-out tests were performed on Ta to determine feature parameters (if applicable) and to test some other classifier conditions, e.g. distance measures, scale-dependency, K value.
- For the final classification performance the hold-out tests were performed using Ta as the training and Tt as the testing set.

In the final set of experiments, the Fisher Linear Discriminant (FLD) and Back Propagation Neural Network (BPNN) classifiers are trained and the evaluation results were compared with KNN. Extended information about these classifiers can be found in various texts [31–33]. See Section 7.4 for a general comparison of these classifiers with KNN.

8.2 Training Set Experiments

8.2.1 Features: Parameters and Performances

The histogram and area granulometry feature both have only one parameter: the number of colours and area set, respectively; the correlogram feature has two parameters: the number of colours and distance set. The shape based features (Hu moments and relative shape measurements vector) have no parameters. The leave-one-out evaluations on Ta are performed for a set of fixed K values ($K = \{1, 3, \ldots, 15\}$) and the results were averaged. See Section 7.4.1.2 for an explanation of K value in KNN classification. The parameters for the features (e.g. number of colours, distance sets) were chosen according to skew insensitive average accuracy performance ($Acc*$).

All feature elements in all feature vectors were independently normalised for zero mean and unit variance. This was performed assuming a normal distribution for each feature element. A normal distribution empirically would suggest that about 99.7% of data lie within 3 standard deviations [166]. However, in the distance metric experiments, it is shown that this procedure is not necessary; it degrades the performance if the relative distance R_{L_1} metric is used. See Section 7.4.1.3 for a description of the simple distance metrics that are used in this study.

8.2.1.1 Histogram

As discussed in Section 7.2, the most important aspect of the histogram feature is the number of the distinct colour bins. Hence, the images are re-quantised to reduce the number of possible colours and to observe the efficiency of the histograms for the classification.

Another aspect of the histogram feature is its scale dependency. The normalisation with respect to total number of pixels provides a scale-invariant feature of only colour distribution (i.e. normalised histogram). However, here the histogram feature counts the occurrence of the colours in the *arbitrary sized* stained objects which are

N_c	8	16	32	64	128	256	512
H	77.5	82.7	**85.1**	82.7	81.6	76.7	70.4
h	66.5	71.9	80.0	78.6	81.5	**82.6**	81.2

Table 8.1: Feature Evaluations: the normalised and un-normalised histogram $Acc*$ performances (in Ta) for different numbers of colour bins.

not necessarily of the same scale. Because of this problem, it is not guaranteed that the normalised histogram will perform better than the un-normalised (pixel count) histogram. Table 8.1 shows the (Acc^*) performances for the different numbers of colours tested separately for normalised and un-normalised histogram features (denoted by h and H, respectively).

The N_c (number of colour bins) values were chosen from the powers of two. Among the tested values $(8, 16, 32, 64, 128, 256, 512)$ 32 demonstrated the best results for un-normalised histogram H; and was used in consecutive comparisons. It can be observed that the performances were gradually lower for the other (higher/lower) bin values. On the other hand, the performances of the normalised histogram (h) were better for the higher number of colour bins (i.e. 128 and 256); however, lower than the 32 bin un-normalised histogram. It can be stated that the un-normalised histogram is more efficient (feature size vs. accuracy). This is due to the arbitrary object size. The normalised histogram removes the clear distinction between large and small objects, which possibly was used as a discriminatory aspect by the un-normalised histogram feature.

As discussed in Section 7.2.6, an alternative way to provide scale-independence is to scale the images prior to feature extraction. Hence, the performance of the un-normalised histogram feature ($N_c = 32$) was tested with the images that were scaled with respect to the average RBC area (calculated by area granulometry of the whole image). The performance dropped from 85.1 to 83.1. Note that this value (83.1) is

greater than the best (256 bin) normalised histogram performance (82.6). Thus, the results suggest that scaling the images with respect to average cell area is better than normalising the histogram using the total number of pixels. However, the degraded performance suggests that the average cell area measure, which is calculated by the area granulometry, is not accurate for image scale normalisation.

Finally, it must be noted that any discrete value of the number of colours can be used for the quantisation; it may be possible to find other values which may be better for the training set (Ta).

8.2.1.2 Correlogram

The details of the correlogram feature were given in Section 7.2.2. The correlogram feature has two parameters: the number of colours (N_c) and the distance set (D). Considering the two parameters might be dependent, the best approach would be to perform a search for the different possible combinations. Table 8.2 shows the measured Acc^* values for several combinations of a limited set of number of colours and distance sets. The distance sets were formed of consecutive integer distance values; the maximum distance was taken as a parameter. For example, the value 12 in Table 8.2 corresponds to the distance set $D = \{1, \ldots, 12\}$. Note that, a correlogram vector of a bigger maximum distance value includes that of smaller.

For a fixed value of D, it can be observed that the effect of the number of bins parameter (N_c) is similar to that of the histogram feature. The best row of Table 8.2 is 32, and the best performance was achieved using $D = 16$. It can be observed that the results were more affected by N_c than the distance parameter D. As D value increases, the calculated feature takes into account more distant colour co-occurrences. Hence, we expect that the optimal distance set would be related to the scale and possibly the size of the objects. However, Table 8.2 shows that there is no regular pattern for the distance set dependency with the exception of the $N_c = 32$ row which shows slight increases with D. A possible explanation can be that the objects being both in

N_c \ D	4	8	12	16
8	79.4	79.3	79.8	79.1
16	79.6	80.2	80.6	80.4
32	82.7	83.2	83.4	**83.8**
64	78.5	79.7	79.4	79.1
128	77.6	78.0	78.0	78.8

Table 8.2: Feature Evaluations: The correlogram feature $Acc*$ performances (in Ta) for different numbers of colours and different distance sets.

arbitrary size and scale, hinders such an optimal value. Thus, the optimal distance set for one scale was not optimal for another. However, according to $N_c = 32$ row the performance could be improved using higher D values. Table 8.3 shows the results of $N_c = 32$ for higher D values (i.e. 20,24, and 32). It can be observed the performance increases up to $D = 24$ value. However, the D parameter and scale dependency was further tested with the images that were scaled with respect to the average RBC area (calculated by area granulometry of whole image as for histogram). Table 8.3 shows the results of these tests ($32_{Ag\,Sc}$). Although, the performances were lower than its counterpart, a regular pattern of the D dependency was observable. The performance made a peak at the $D = 12$ value.

Normally, we should choose the $N_c = 32$, $D = 24$ for the following experiments. However, the computational cost of extracting and comparing $N_c = 32$, $D = 24$ correlogram is significantly higher than $N_c = 32$, $D = 12$. Considering the performance increase is not significant, also lower than that of the histogram feature, we set the parameters to $N_c = 32$, $D = 12$. This corresponds to a correlogram of total 384 feature elements.

D / N_c	4	8	12	16	20	24	32
$32_{Un\,Sc}$	82.7	83.2	83.4	83.8	84.1	**84.3**	84.2
$32_{Ag\,Sc}$	79.7	80.1	**80.7**	80.2	79.8	79.7	78.8

Table 8.3: Feature Evaluations - correlogram 2: The correlogram feature *Acc** performances (in Ta) for different numbers of colours and different distance sets. The first row shows ($32_{Un\,Sc}$) the results when the images were not scaled. The second row ($32_{Ag\,Sc}$) denotes the performances obtained using the pre-scaled images with respect to average cell area.

Radius	5	10	15	20
Un Sc	75.6	80	**81.4**	79.7
Ag Sc	73.3	73.3	74.2	66.4

Table 8.4: Feature Evaluations: The area granulometry feature *Acc** performance (in Ta) for different area sets. The first row shows ($_{Un\,Sc}$) the results when the images were not scaled. The second row ($_{Ag\,Sc}$) denotes the performances obtained using the pre-scaled images with respect to average cell area.

8.2.1.3 Area Granulometry

The area granulometry feature has a single parameter: the area set Λ. The parameter is scale dependent. The area sets were formed of consecutive integer radius values and the maximum radius was taken as the parameter. We have tested a limited set of values: (5,10,15,20). Table 8.4 shows the calculated *Acc** values. The best performance was obtained by the value 15 which corresponds to an area set of $\Lambda = \pi\{1^2, \ldots, 15^2\}$; and thus a feature vector of 45 elements (15 for each colour channel).

As for the correlogram, we tested to see whether the performance could be improved using the pre-scaled images with respect to the image average RBC area. Again there was no improvement; the performance dropped from 81.4 to 74.2 (see $_{Ag\,Sc}$ in Table

M	S	Rnd
55.1	63.0	49.9

Table 8.5: Feature Evaluations: The $Acc*$ performances (in Ta) of the Hu moments (H), relative shape measurements (S), and random (Rnd) features.

8.4). We set the maximum radius parameter to 15 for the following experiments.

8.2.1.4 Hu moments and Relative Shape Measurements

The two shape based features have no parameters. The average results of the tests on Ta are shown in Table 8.5. It can be observed that both feature performances were lower compared to the previously tested three features. In order to provide a comparison, a random feature of 5 elements was generated and used in the same tests. $Acc*$ of the random feature was as expected: ~50%. It can be seen that the performance of the Hu moments feature M is slightly better than the random feature's; the relative shape measurements feature (S) is relatively higher. However, if compared to the colour based features they were significantly low. This may be related to two different things: the shape of the objects (as in binary form) may not be discriminative; the extracted stained binary objects may be too irregular or too noisy for the limited discriminative power of features.

8.2.2 Feature Combinations

In this stage, we would like to investigate whether the performance can be improved by combining the features. The previous tests (above) showed that the feature success order (according to the Acc^*) is H, C, G, S, M. Table 8.6 summarises the results according to Acc^* but also shows the measured Acc values to observe if there is any trade-off which may hinder the best feature conclusion. It can be observed that there is a trade-off indeed between the Acc and $Acc*$ results of C and H. It can be concluded

	H	C	G	S	M
Acc	93.0	93.1	89.8	84.2	83.1
$Acc*$	85.1	84.3	81.4	63.0	55.1

Table 8.6: Feature Evaluations: Summary of the Acc and $Acc*$ performances (in Ta) of the individual features.

	$CHGS$	HGS	CGS	HG	CG	HS	CS	GS	$SFFS^{(24)}$
Acc	**93.86**	92.98	93.81	92.82	93.53	92.60	93.35	89.41	92.41
$Acc*$	86.22	**86.54**	86.21	86.40	85.57	84.65	84.03	81.48	86.14

Table 8.7: Feature Evaluations: Training set (Ta) performances (Acc and $Acc*$ of the concatenated features and the SFFS reduced $CHGS$ feature subset.

that both features C and H perform well; however the former favours the negative class more than the latter. Nevertheless, if the additional computational complexity of the correlogram feature is considered, the histogram feature is highly efficient and preferable.

The moment feature M was discarded from the remainder of the experiments since its performance was quite close to the random feature. Using the remaining four features, eight different concatenations were formed: $C + H + G + S$, $H + G + S$, $C+G+S$, $H+G$, $C+G$, $H+S$, $C+S$, $G+S$. The numbers of feature elements in the concatenated feature vectors were as follows: $C+H+G+S = 384+32+45+6 = 467$, $H+G+S : 32+45+6 = 83$, $C+G+S = 384+45+6 = 435$, and so forth. The first eight columns of Table 8.7 show the results of the different tested concatenations. It can be observed that the concatenations performed slightly better than their individual features; HGS and CGS features inherited the previous trade-off of H and C features for the different accuracy measures. The concatenation of all features ($CHGS$) did not perform significantly better than HGS and CGS features.

Note that the every single feature of the feature vectors were independently nor-

malised and had values in range $[0-1]$. Therefore, the impacts on the distance calculations were equally weighted for every element in the concatenated feature vectors. However, the direct concatenation of a large feature vector with a small feature vector produces results more affected by the large vector. For example, the *CHGS* results not being significantly better than *CGS* results can be related to this; also to the possibility that the C and H features being correlated.

As discussed in Section 7.3, concatenation is not an efficient approach to obtain an improved set of features since many features might be redundant or correlated. Hence, we formed a feature pool using all features in *CHGS* concatenation. Then we used the SFFS (sequential floating forward selection [135]) algorithm to investigate whether performance can be improved. Here the KNN misclassification percentage was used as a criterion (i.e. wrapper approach) for SFFS. However, due to computation time considerations only the positive examples in Ta were evaluated with the leave-one-out approach for SFFS. Thus, the SFSS process on *CHGS* tried to maximise only the positive class accuracy (i.e. true detection rate, sensitivity). The K value for KNN and the desired final feature size was set to 5, 36 respectively (K value dependence is investigated later). Figure 8.1 shows the plots of the leave-one-out *Acc* and *Acc** results with respect to the increasing feature size. It can be observed that for several values of feature size (i.e. >13) the *Acc** value is higher than the full set of features having total 467 features. The maximum *Acc** value was achieved by a subset having 24 features; however it can be seen that after 15 features the *Acc* was not significantly affected by the increasing feature number. This is because of the criterion used in SFFS, which tried to maximise the positive class accuracy (i.e. sensitivity). Hence, the features selected after this point increases the sensitivity but reduces the negative class' accuracy (i.e. specificity). However, these results were obtained for a specific K value (i.e. 5). The averaged *Acc* and *Acc** values over the set of the previously used set of K values (i.e. $\{1, 3, \ldots, 15\}$) as done for the other features are shown in Table 8.7. It can be observed that the *Acc* and *Acc** performances were only slightly below of

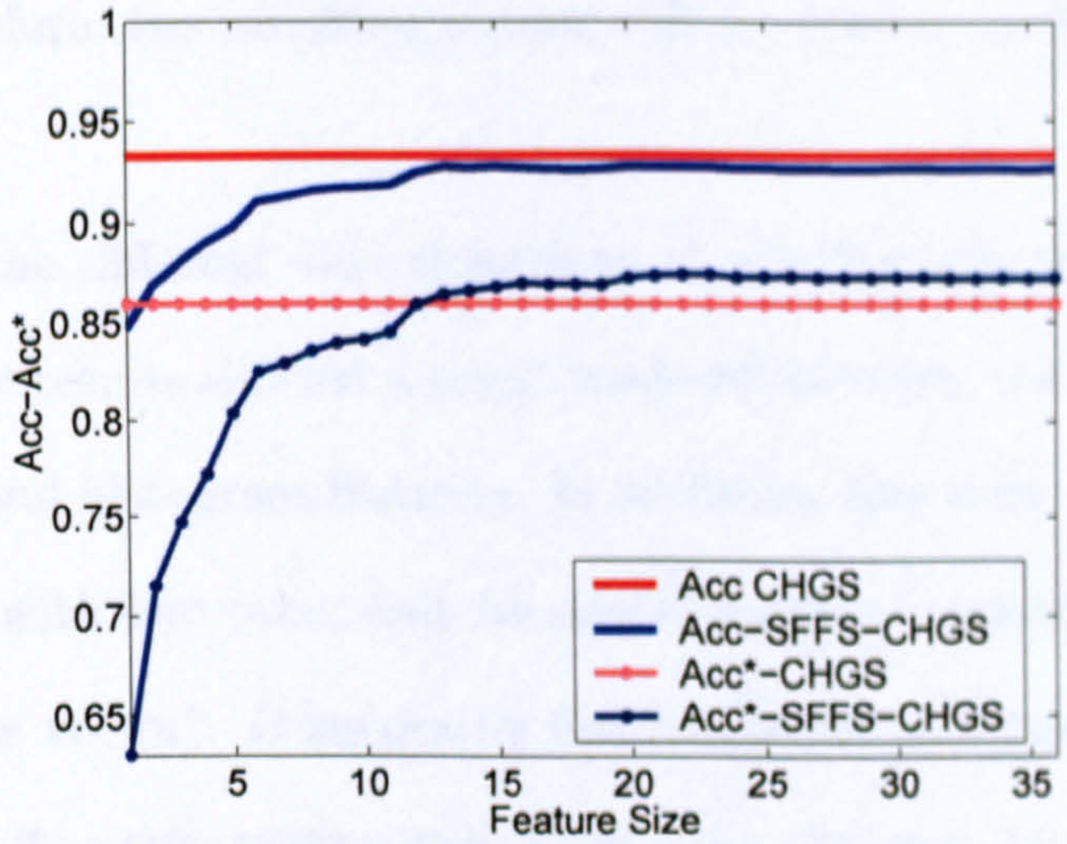


Figure 8.1: Feature Evaluations - SFFS performance vs. feature size: *Acc* and *Acc** versus SFFS selected feature subset sizes

the most successful concatenation performances although only 24 features were used. Moreover, if compared to the individual feature vector performances (see Table 8.6), it can be seen that the SFFS calculated subset achieved higher *Acc** than all feature vectors. The subset of 24 features included 12, 6, 5, and 1 feature elements from the C, H, G, S feature vectors respectively.

However, it must be noted that the results shown here do not guarantee a generalised solution (e.g. a good performance in the test set). Since the KNN accuracy on the training set is used as a criterion for SFFS, the results might be too optimistic. This can be simply caused by the over-fitting of the selected subset to the training samples. The over-fitting is the case where an estimate of, say, a function is very closely fitted to reduce the mean error (i.e. bias) whereas it results in a high variance and as a consequence a poor generalisation. There are practical solutions for over-fitting, such as validating the current estimate on a separate set to continue or stop training. However, this relies on a monotonic learning scheme, which does not exist in the KNN wrapped SFFS. Alternatively, the different subsets calculated by SFFS can be tested on a separate validation set; and the best one can be selected for the tests. However, this requires a further separation of the training or test set, which complicates our

comparisons. Therefore, the resulting subset will be tested on the test samples T_t to reveal its success.

To summarise, the different concatenations of the features were evaluated on the training set Ta. The results showed a small trade-off between the Acc and $Acc*$ values of the correlogram and histogram features. In addition, this trade-off was observable in the concatenations with two other less successful features (granulometry and relative shape measurements vector). Considering the computational complexity and the size of the correlogram, its performance was below expectations. On the other hand, the SFFS (sequential floating forward selection [135]) algorithm is employed to select a reduced subset of features using the pool of all four features. The resulting subset was formed by 24 features, which produced comparable results to feature concatenations on the training set. The following experiments utilise the HGS feature and the subset of features which was selected by the SFFS algorithm.

8.2.3 Feature Normalisation and Distance Metrics

A discussion about the different metrics can be found in Section 7.4.1.3. The results shown up to this point used KNN with the relative distance metric (RL_1). The performance of the selected feature concatenation HGS was assessed once again using the different distance metrics in order to see the effect of the distance metrics on the classification. The first two rows of Table 8.8 show the averaged Acc and $Acc*$ (on Ta) results when the features were normalised to have zero mean and unit variance. It can be seen that the performance of L_2 Euclidean distance is the worst of the three; whereas the difference between L_1 city block and RL_1 relative city block distance is marginal. It can be seen that for the un-normalised HGS feature the difference between RL_1 and the other is significant. Moreover, compared to the normalised features the RL_1 performance was better.

RL_1 is more sensitive to changes between the smaller feature values (see Figure 7.5).

norm	L_2	L_1	RL_1
Acc	91.95	92.90	92.98
*Acc**	84.17	86.05	86.54

unnorm	L_2	L_1	RL_1
Acc	90.12	91.06	91.09
*Acc**	80.18	81.38	87.50

Table 8.8: Feature Evaluations: feature normalisation and different distance metrics: The normalised and un-normalised *HGS* feature's *Acc* and *Acc** performances for the different metrics.

The normalisation process compresses all feature values to $[0, 1]$ range. The results showed that the feature normalisation degrades the RL_1 performance. However, the same was observed for the other distance metrics. Hence, this implies that the features, which have smaller values, are more relevant. If we consider the biggest values, for example for the histogram feature, would be due to cell background colour.

Motivated by these results, the SFFS algorithm was run once more using the un-normalised features using the same settings as before. As expected the *Acc*, *Acc** performances were higher than for the normalised feature case: 93.4, 87.4 respectively (compare to Table 8.7); however, still slightly below that of the *HGS* feature (Table 8.8). The newly selected feature subset size was 33; included 9, 5, 18, 1 features from *C*, *H*, *G*, *S* features respectively. It was surprising to see that 18 features of *G* were included in the best subset given its peer individual performance.

The following evaluations use un-normalised features with the RL_1 distance metric.

8.2.4 K value Dependence and Unbalanced Training Data

It can be noticed that the average accuracy (*Acc*) results shown so far are all above 85%. Also, it can be seen that *Acc* values of the SFFS selection accuracy in Figure 8.1 starts with (Feature Size=1) ~85%. In fact, for *Ta*, the *Acc* = 84.17 value corresponds to classifying every object as non-parasite. This is simply the effect of the unbalanced (i.e. skewed, unequally distributed) training set. Since this was an expected condition the

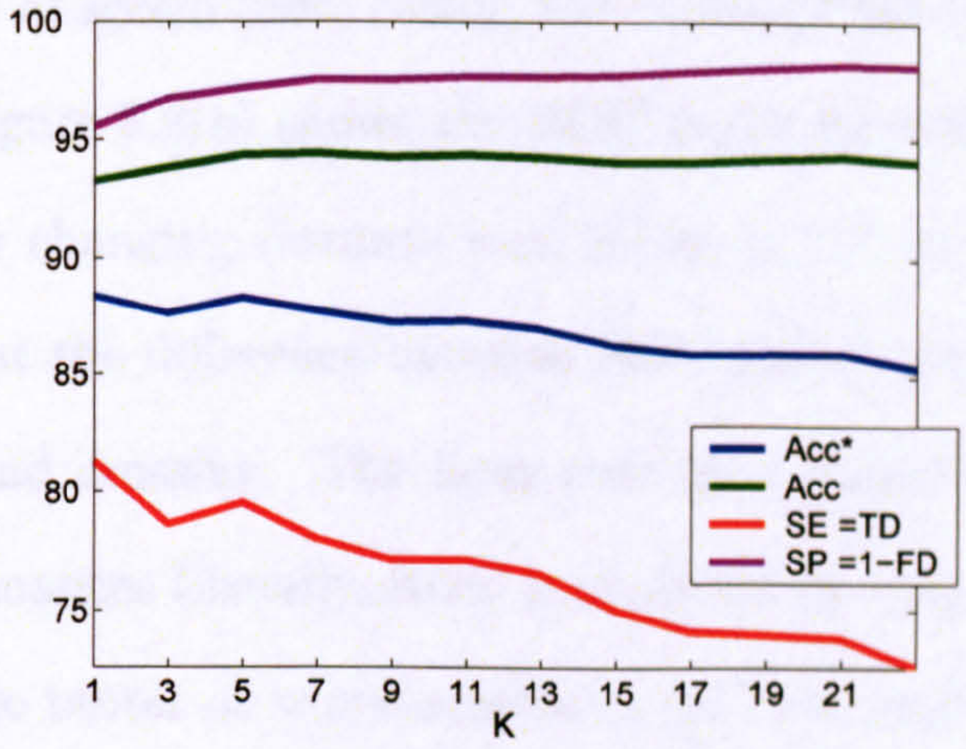
alternative measure $Acc*$ was used together to interpret the results and make choices among several different conditions. However, in KNN classification and most other classifiers the decision is biased towards the majority class because the most populated class is favoured to the other one. For KNN, it can be expected that this bias would be higher for higher K values.

Until now, we have considered the accuracy measures which were averaged for a set of K values. For the current discussion it will be useful to observe the dependence of the different accuracy results on the K values. Figure 8.2 shows the Acc, $Acc*$, SE, SP values calculated for different K values for the $IIGS$ feature. It can be observed that although sensitivity SE (i.e. true detection) rate decreases (when $K > 5$) with increasing K values, the specificity SP (1-false detection rate) increases slowly. As a consequence the skew insensitive accuracy ($Acc* = SE + SP$) measure is slowly decreasing whereas the accuracy is relatively stable. The $K = 1$ value provides the highest sensitivity (81.3); the $K = 5$ value is the second best (79.4) with better specificity SP (and accuracy Acc). Thus, for any K value it is not possible to obtain a sensitivity higher than (81.3) which is a direct cause of the unbalanced training set.

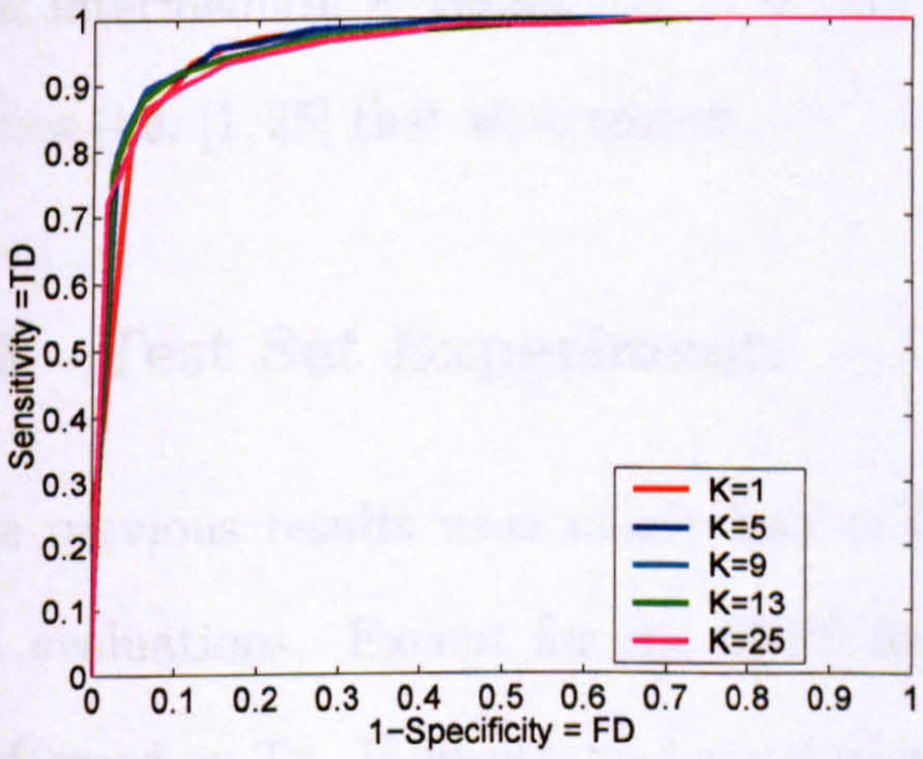
There are several studies concerned about unbalanced learning or dealing with skewed data [167]. The methods include under/over sampling majority/minority class samples; varying the decision thresholds; incorporating misclassification costs. In practice, they all provide biased classification schemes, which are more conservative for the minority class. We followed the approach in [168] which modified the neighbour distance calculation to obtain a biased KNN classifier: Let x_i^c show a training sample of the class c, the biased distance calculation for a query q can be written as follows:

$$D_b(q, x_i^c) = D(q, x_i^c) * \tau_c \tag{8.1}$$

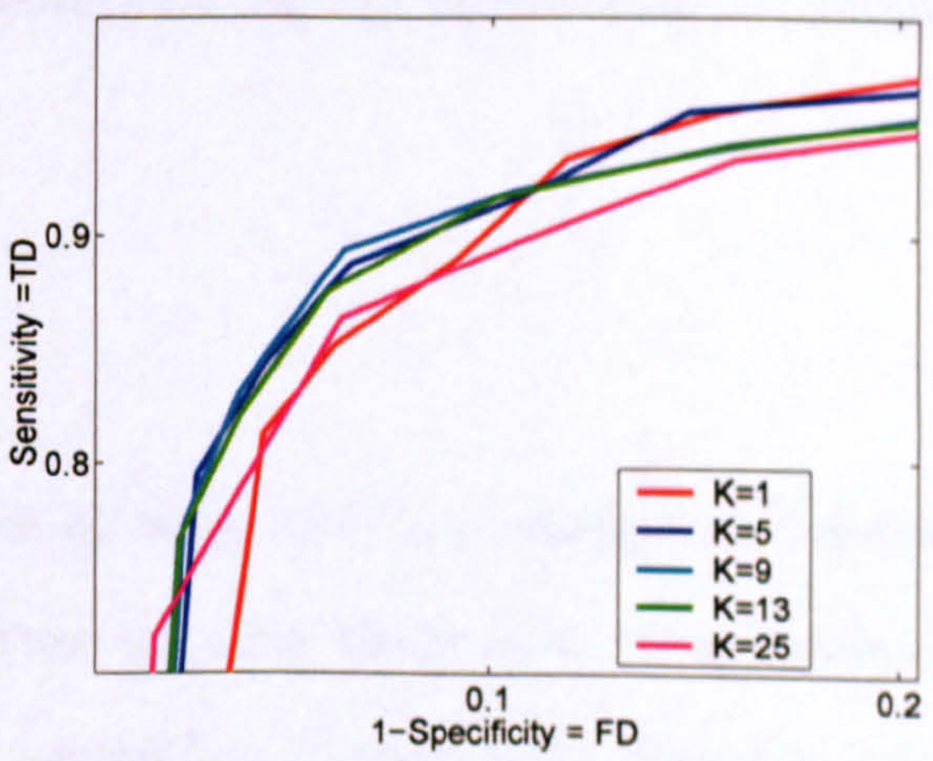
where τ_c represents the class specific distance weighting constant. Setting τ_c to a value smaller than 1 puts a bias in favour of the class c. This is a simple trick to break the unbalanced data bias, and to allow KNN to perform on accuracy values other than the



(a)



(b)



(c)

Figure 8.2: The K value dependence and ROC analysis of the KNN with the unbalanced training set: (a) The dependence of accuracy results on K value. (b) ROC curves for different K values, (c) [0-0.2,0.7-1.0] regions of the ROC curves.

one that is dominated by majority class accuracy.

In this way, using the biased KNN classifier detection can be performed in more preferred sensitivity-specificity trade-offs. Moreover, as previously implemented for stained pixel detection, ROC (Receiver Operating Characteristics) curves can be plotted and results can be analysed more easily; the choice of the operating character can be left to the user. Figure 8.2(b) shows the ROC curve for several different K values (i.e. $[1, 5, 9, 13, 25]$) for changing distance bias values $0 < \tau_c < 10$ (for parasite class). It can be observed that the difference between ROC curves for the different K values were not significant and crossing. The later condition makes it difficult to compare the different K performances (usually, ROC is analysed by area under the curve). The different values provide better or worse results in different regions of the ROC space. However, using a close up look to upper left corner (Figure 8.2(c)), we can conclude that intermediate K values (i.e. $[5, 9, 13]$) performed slightly better than the marginal values (i.e. $[1, 25]$ that were tested.

8.3 Test Set Experiments

The previous results used nearly half of the all data (Ta) and performed leave-one-out evaluations. Except for the SFFS feature selection there were no optimisations performed on Ta. However, we have chosen several feature parameters based on the Ta $Acc*$ performances and performed several tests on the same data. Hence, the selected conditions had to be once again tested on the remaining data (hold-out evaluation) to reveal the generalisation performance. The hold-out evaluation configuration can be summarised as follows: the Ta set was used for training whereas the Tt set was used for testing; K value of the KNN was set to 5; RL_1 distance metric was used; the data was not normalised (for zero mean and unit variance); KNN denotes the basic KNN classifier (i.e. without bias). Here, instead of the skew insensitive accuracy measure ($Acc* = (SE + SP)/2$) the sensitivity and specificity measures were reported

separately together with the average accuracy (Acc) and positive/negative prediction values (PPV/NPV).

Table 8.9 shows the results of HGS and $SFFS$ which selected from $CHGS$ a subset of 33 features. KNN was used with the same settings for both features. Additionally, Table 8.9 shows two other classifier results acting on the selected features. These classifiers were the Fisher Linear Discriminant (FLD) [33] and the Back Propagation Neural Network (BPNN). For FLD we have used the implementation of the Statistical Pattern Recognition Toolbox for Matlab [169]. In order to build and train a 2-layer BPNN we have used the Neural Network Toolbox (Version 4.0) for Matlab distributed by Mathworks. BPNN for HGS and $SFFS$ were both constructed with 8-4-1 number of neurons in respective layers (i.e. input-hidden-output). The HGS network had 692 total weights whereas the SFFS network had 300. The employed learning algorithm was gradient-descent with momentum [32]. The learning rate and momentum constants were set to 0.05 and 0.95, respectively. It was not possible to train the BPNN networks using un-normalised features; hence, the features were normalised to have zero mean and unit variance assuming a normal distribution. The BPNN networks were trained several times (25) with different (random) initialisations; and the mean and standard deviation of the results were provided. The batch training iterations were stopped by cross validations using 10% of the training data. Table 8.9 shows the results for different measures. If we compare the accuracy measures, it can be seen that for both HGS and $SFSS$ the KNN classifier performed slightly better than the two other classifiers. It can be observed that all three classifiers reflected the effects of unbalanced data to the results: although the accuracies (Acc) were high, they were dominated by the non-parasite class accuracy (SP). However, all feature selections were based on the KNN results of Ta. Hence, it is not appropriate to conclude that the KNN performs better than the other classifiers for the parasite detection task. We also tested other features and concatenations with FLD and BPNN. The classifier accuracy order (KNN-BPNN-FLD) was not much different; however, we observed that both FLD and BPNN

Classifier-Feature	Acc	SE	SP	PPV	NPV
KNN-HGS	93.17	72.07	97.45	85.11	94.52
KNN-SFFS	93.02	69.97	97.69	85.98	94.14
FLD-HGS	90.34	73.87	93.68	70.29	94.66
FLD-SFFS	88.68	69.07	92.64	65.53	93.67
BPNN-HGS	92.21±0.4	70.45±1.7	96.62±0.4	80.88±1.9	94.17±0.3
BPNN-SFFS	91.20±0.3	63.63±1.9	96.78±0.2	80.03±1.0	92.90±0.3

Table 8.9: Test set results for different classifiers and features. KNN: K Nearest Neighbour, FLD: Fisher Linear Discriminant, NN: Back Propagation Neural Network, HGS: Histogram, Area Granulometry and Shape measurements feature concatenation, SFFS: Sequential Floating Forward Algorithm selected 33 features from the pool of Correlogram, Histogram, Granulometry and Shape measurements vector feature.

performed slightly better with the *CHGS* feature. It may be possible to tune BPNN to achieve even better results than KNN; however the results imply that the difference would not be significant. The comparison of a complex classifier like BPNN, which depends on many parameters, is complicated task, and beyond the scope of our study.

The primary aim of the SFFS experiments in this study was to see whether a feature selection algorithm could improve the results by reducing the dimensionality of the feature space. The SFSS selected subset of *CHGS* did not perform better than the concatenated feature *HGS*. This can be related to the implementation where we have limited the maximum number of features to 36. However, as can be expected the reduced subset performed best in the KNN (obviously because KNN was used as a criterion). Moreover, it can be concluded the feature size was effectively reduced without significant accuracy trade-off. The scheme can be useful if the speed of the classification is a major concern.

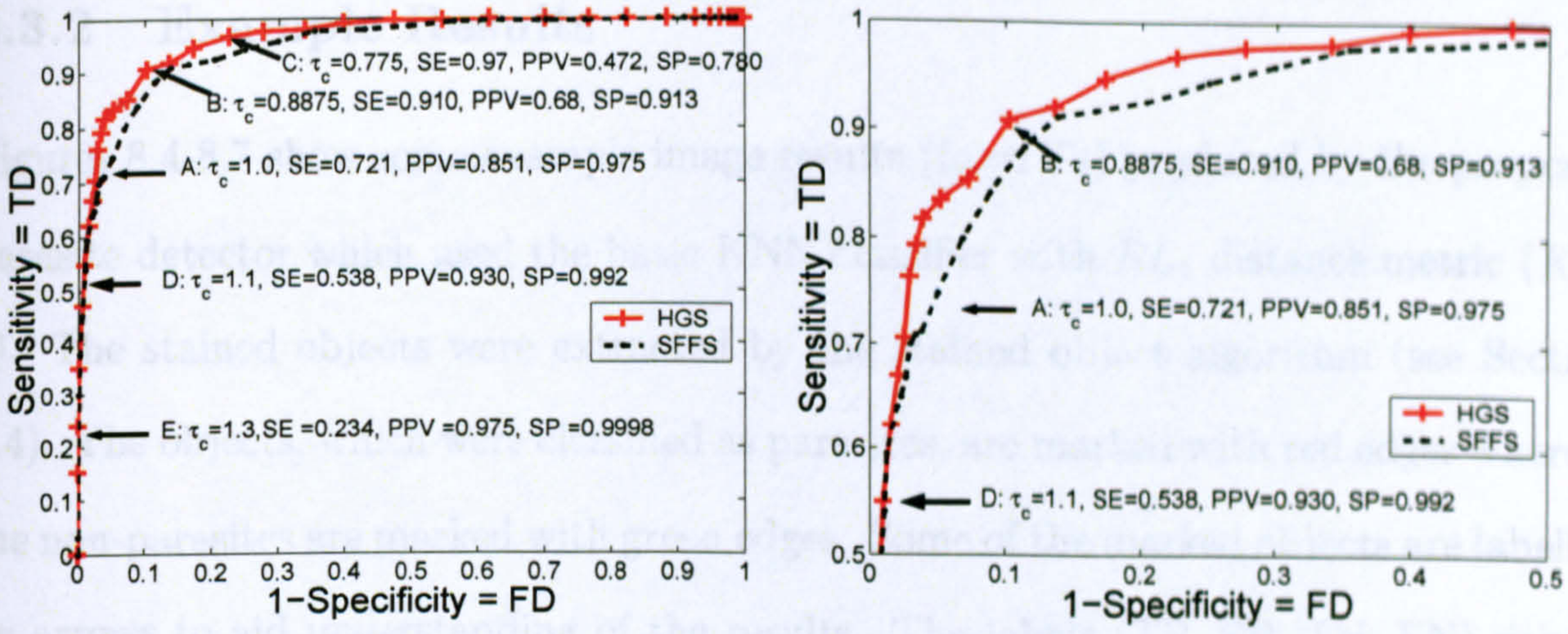


Figure 8.3: Biased KNN classifier evaluated on the test set Tt - (a) whole ROC curve, (b) [0.5-1.0,0.5-1.0] interval of (a).

8.3.1 Biased KNN

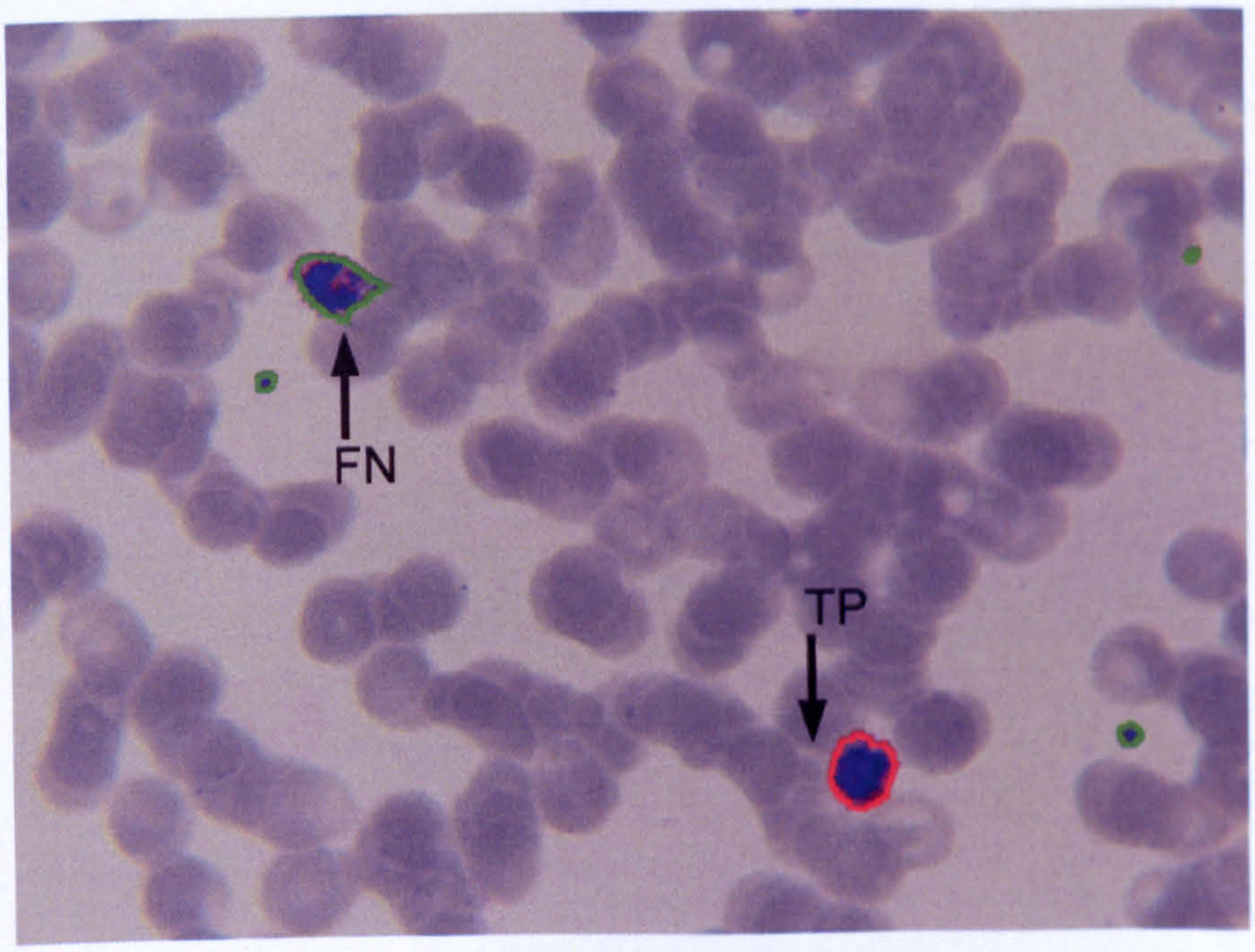
In the previous section, it was shown that the parasite/non-parasite class accuracy unbalance that is due to the unbalanced training set could be relieved when a bias term is incorporated into the KNN distance calculation. This experiment was once again repeated for the test set evaluation. Figure 8.3 shows the ROC curve for biased RL_1 distance function (the parasite class bias was varied between $0 < \tau_c < 10$). It can be observed that by varying the τ_c it is possible to operate KNN in different sensitivity-specificity (SE-SP) trade-offs. The operating point of the basic classifier is marked on the curve (Figure 8.3(a)). By decreasing the τ value for the parasite class higher sensitivity values can be achieved; however with reduced specificity and positive prediction values. For example, if 73% percent specificity (SP) and 43% percent PPV are acceptable, the classifier can provide 97% sensitivity. Thus, flexible detection sensitivity can be provided.

8.3.2 Example Results

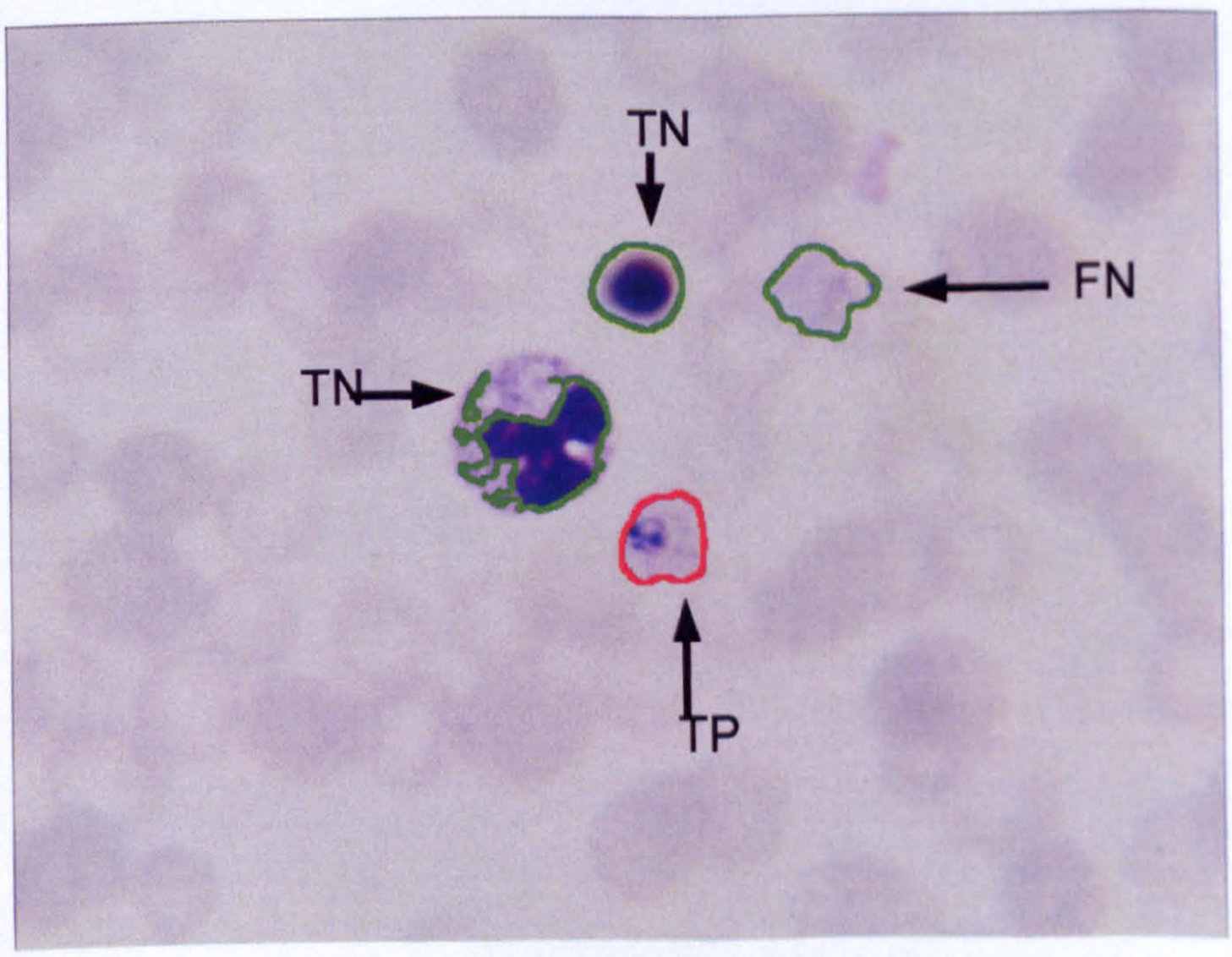
Figures 8.4-8.7 show some example image results (from Tt) produced by the proposed parasite detector which used the basic KNN classifier with RL_1 distance metric ($K = 5$). The stained objects were extracted by the stained object algorithm (see Section 6.4). The objects, which were classified as parasites, are marked with red edges whereas the non-parasites are marked with green edges. Some of the marked objects are labelled by arrows to aid understanding of the results. The labels (TP, FP, TN, FN) denote true positive, false positive, true negative, false negative, respectively (see Section 7.5 for an explanation of the terms). The images are from dense fields, light fields, fields with different colour character and different scales. They include examples of P. Vivax (Figures 8.5(a), 8.5(b) and 8.6(a)), P. Ovale (Figure 8.4(a)), P. Falciparum (Figures 8.4(b) and 8.6(b) and 8.7(a)), P. Malariae (Figure 8.7(b)). Overall, it can be observed that the classifier is able to detect parasites and reject non-parasites. The errors were mostly due to small parasites (i.e. rings) being classified as non-parasites. We will analyse the classification errors in more detail in Section 9.1.2.

8.3.3 Scale-Invariance

There is one more issue to point here. Although the training and test sets contained independent examples from different images, they could belong to the same blood slide. In order to test the generalisation capability of the classifier beyond scope of our training and test sets, we have tested the classifier on some images, which were not included in the UoW slide set. Figure 8.8(a) shows an example P. Vivax image collected from the internet [10]; Figure 8.8(b) shows an example P. Falciparum image from the set of images provided by NIMR, London, UK (see Chapter 3). In the first image one of the parasites was missed, which can also occur in the test set. However, it can be observed that most of the parasites of the second image were missed. It was, however, possible to detect these parasites using the biased KNN classifier. However,

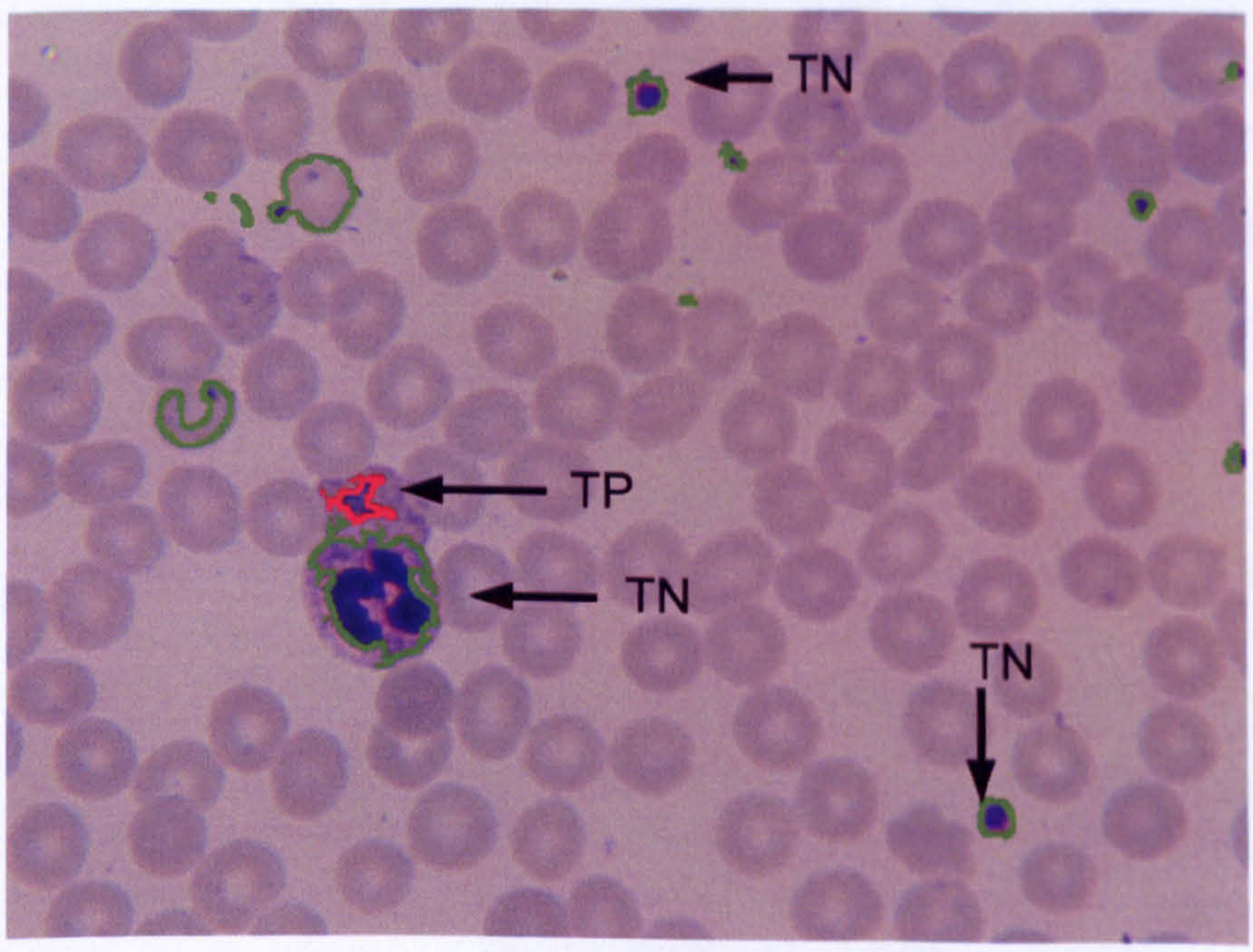


(a)

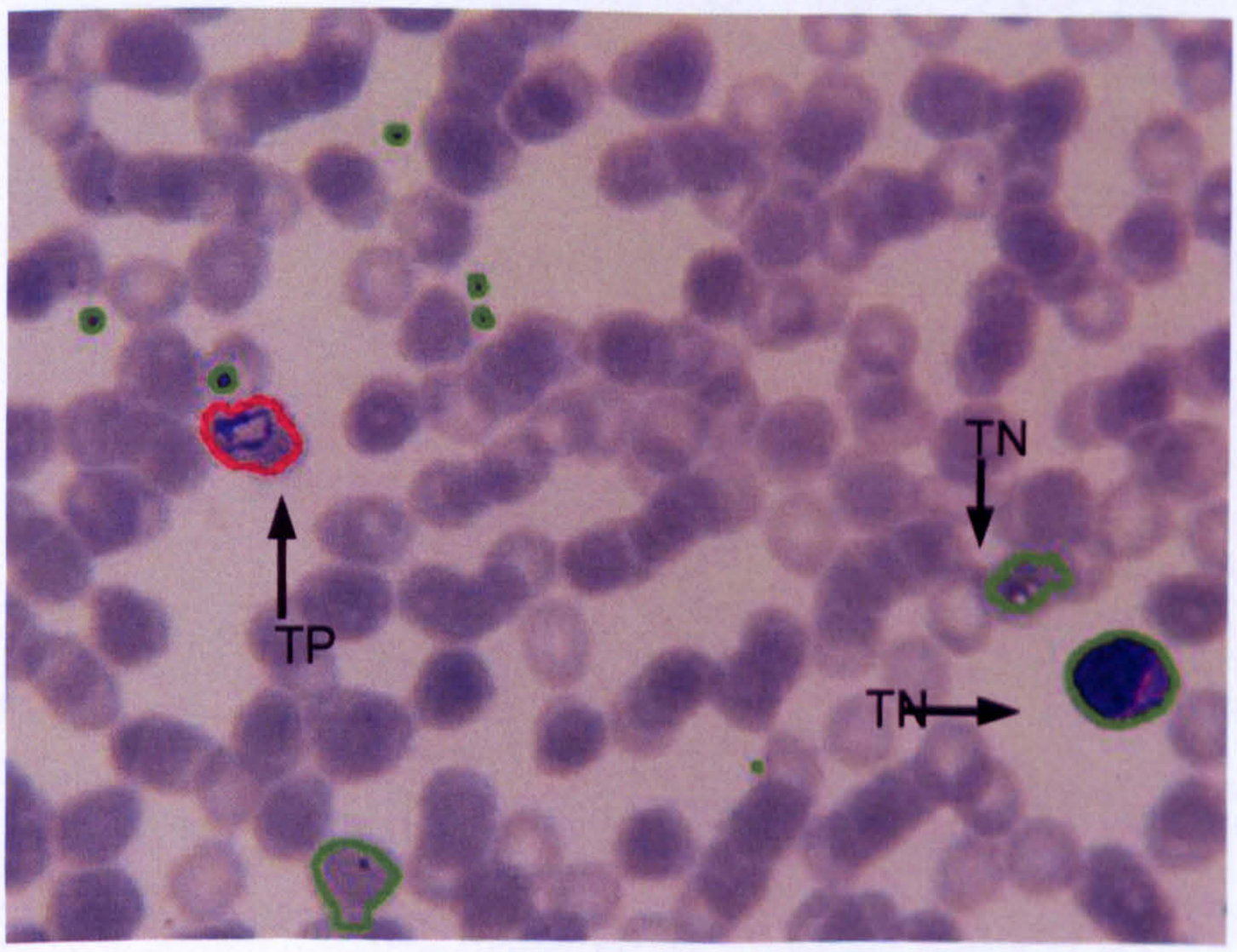


(b)

Figure 8.4: Parasite detection example images 1 from *Tt*: (a) P. Ovale, (b) P. Falciparum. The labels (TP, FP, TN, FN) denote true positive, false positive, true negative, false negative, respectively.

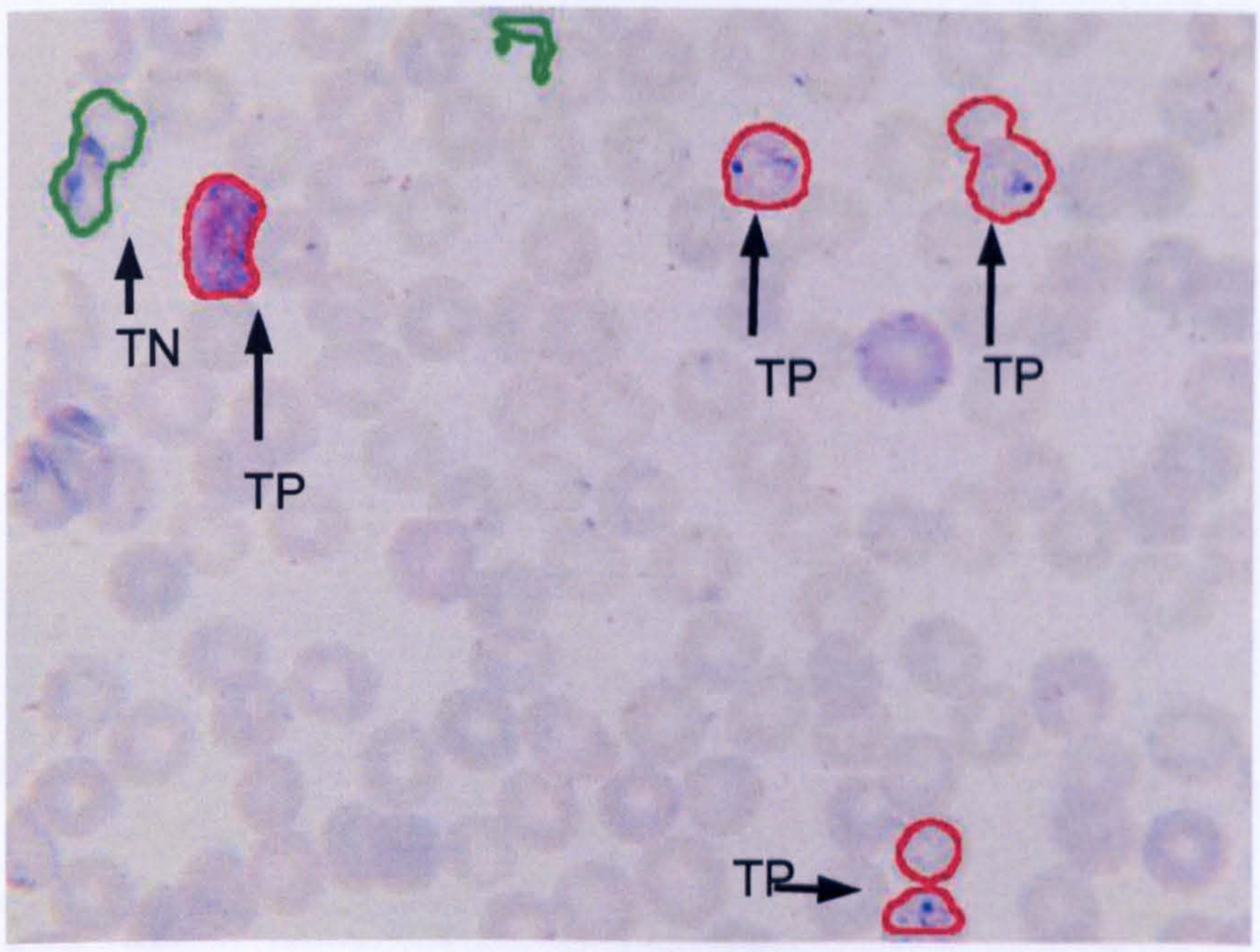


(a)

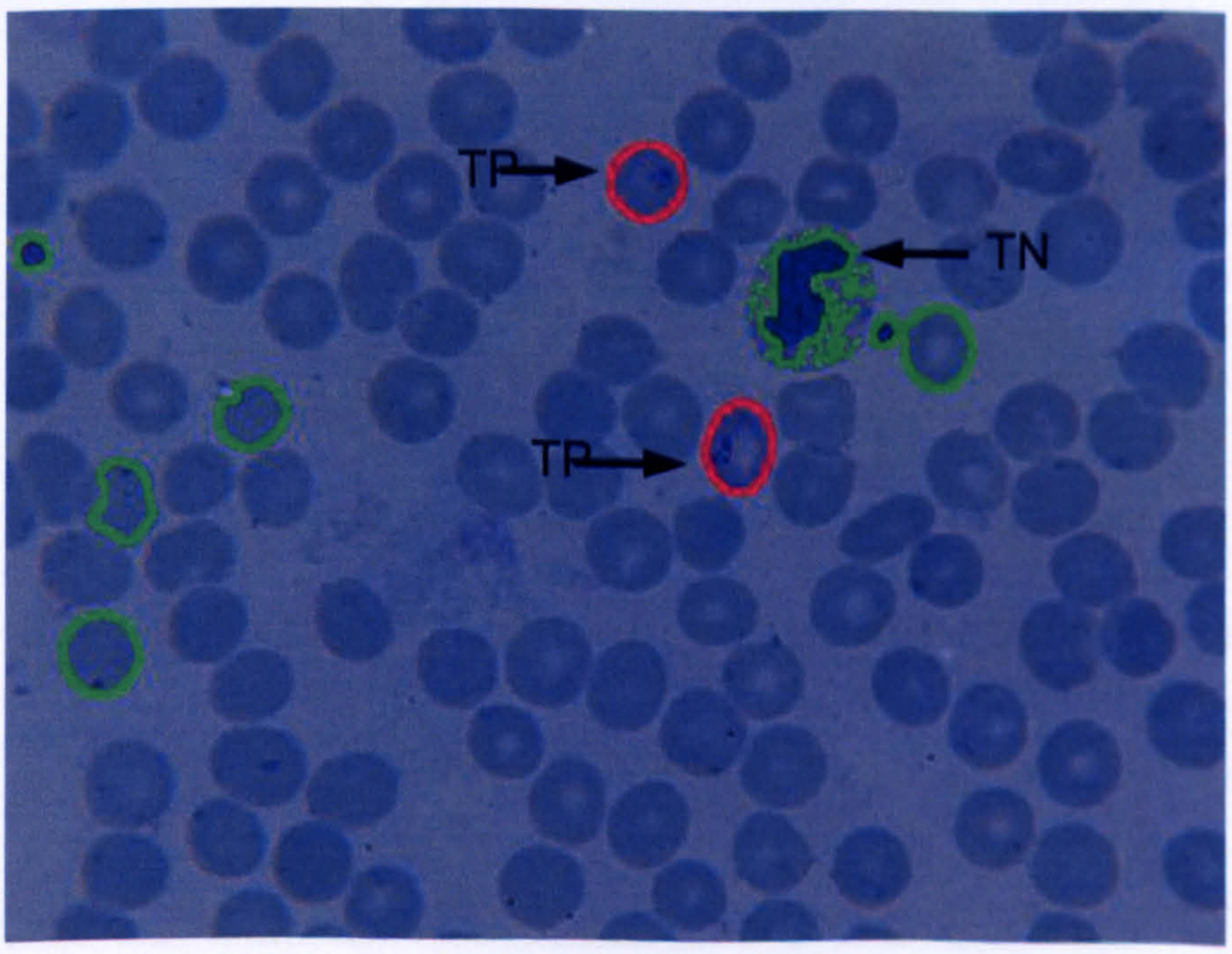


(b)

Figure 8.5: Parasite detection example images from *Tt*: (a-b) P. Vivax. The labels (TP, FP, TN, FN) denote true positive, false positive, true negative, false negative, respectively.

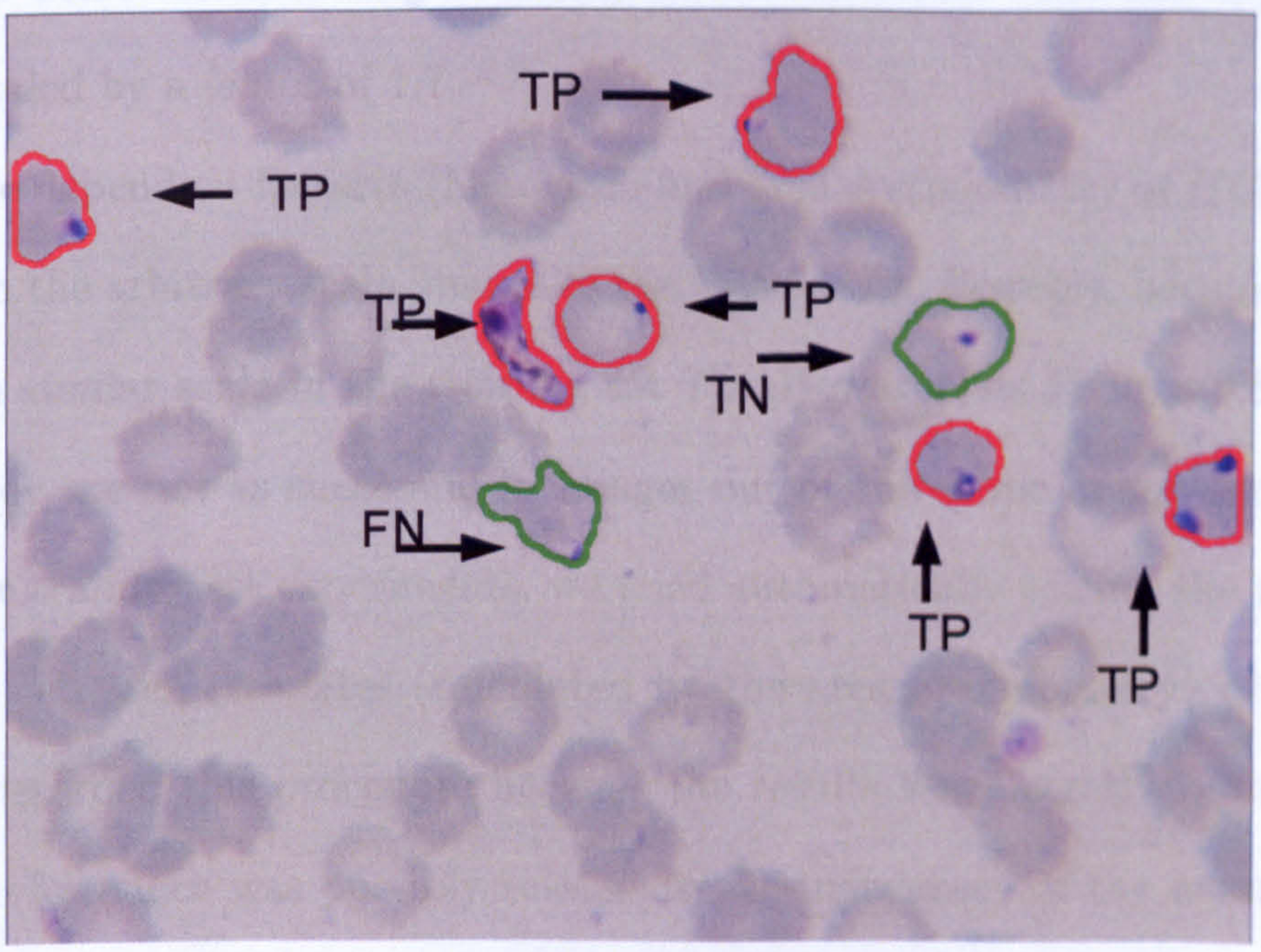


(a)

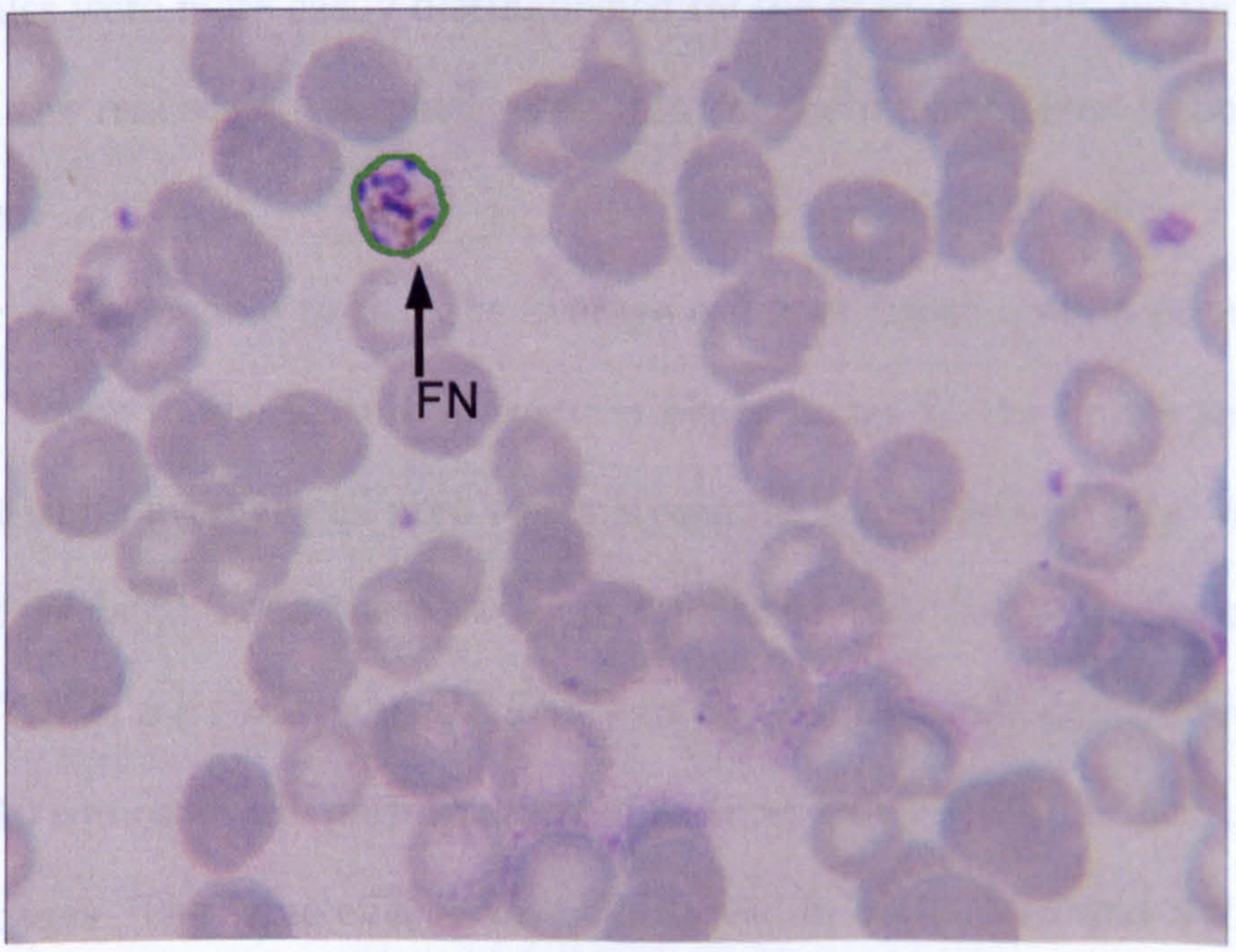


(b)

Figure 8.6: Parasite detection example images from *Tt*: (a) P. Vivax, (b) P. Falciparum. The labels (TP, FP, TN, FN) denote true positive, false positive, true negative, false negative, respectively.



(a)



(b)

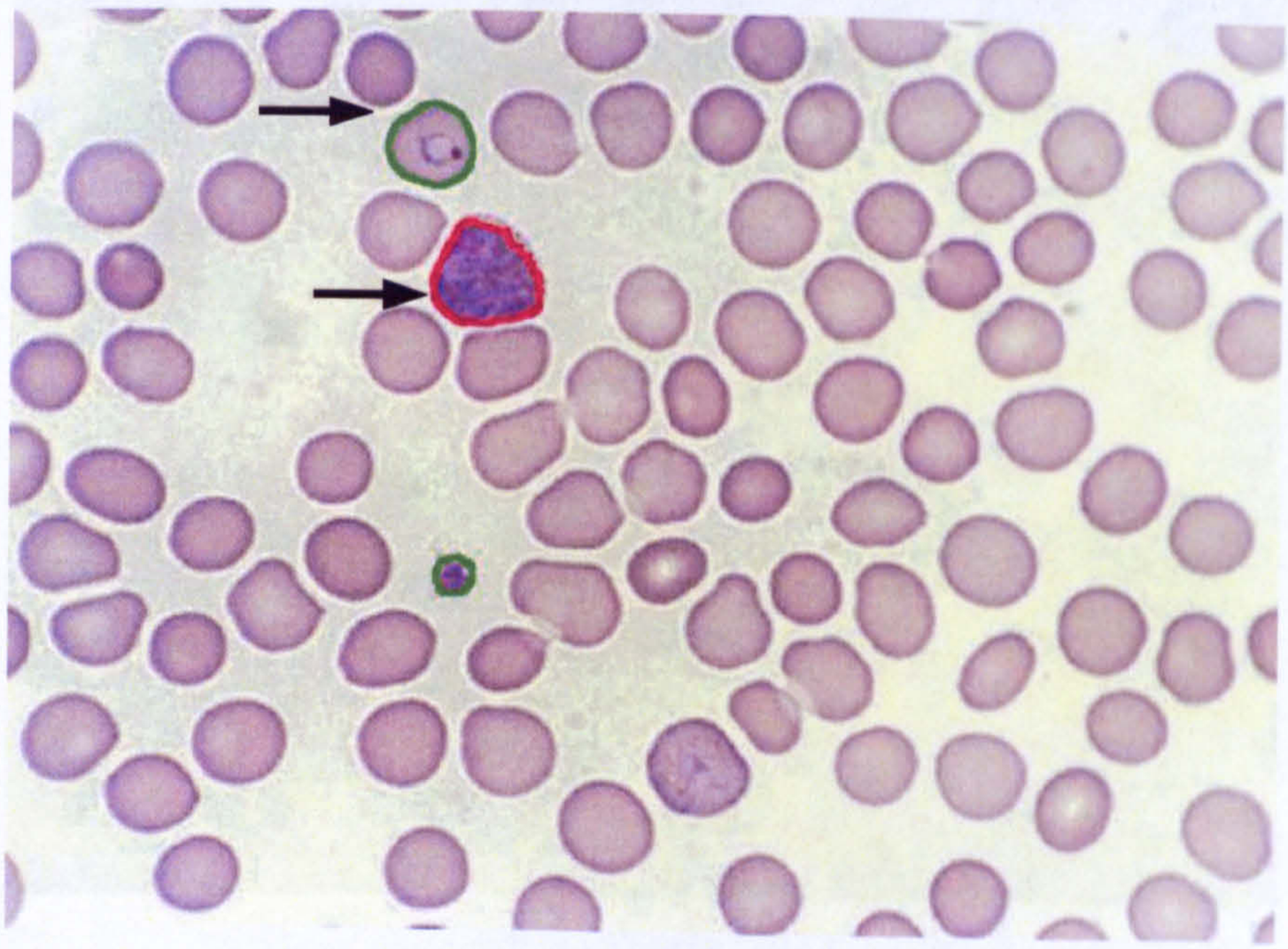
Figure 8.7: Parasite detection example images from *Tt*: (a) P. Falciparum, (b) P. Malariae. The labels (TP, FP, TN, FN) denote true positive, false positive, true negative, false negative, respectively.

the negative results were related to another issue more strongly than just a simple bias setting: scale variance. Figure 8.9(a) shows the result when the input image is manually scaled by a factor of 1.7.

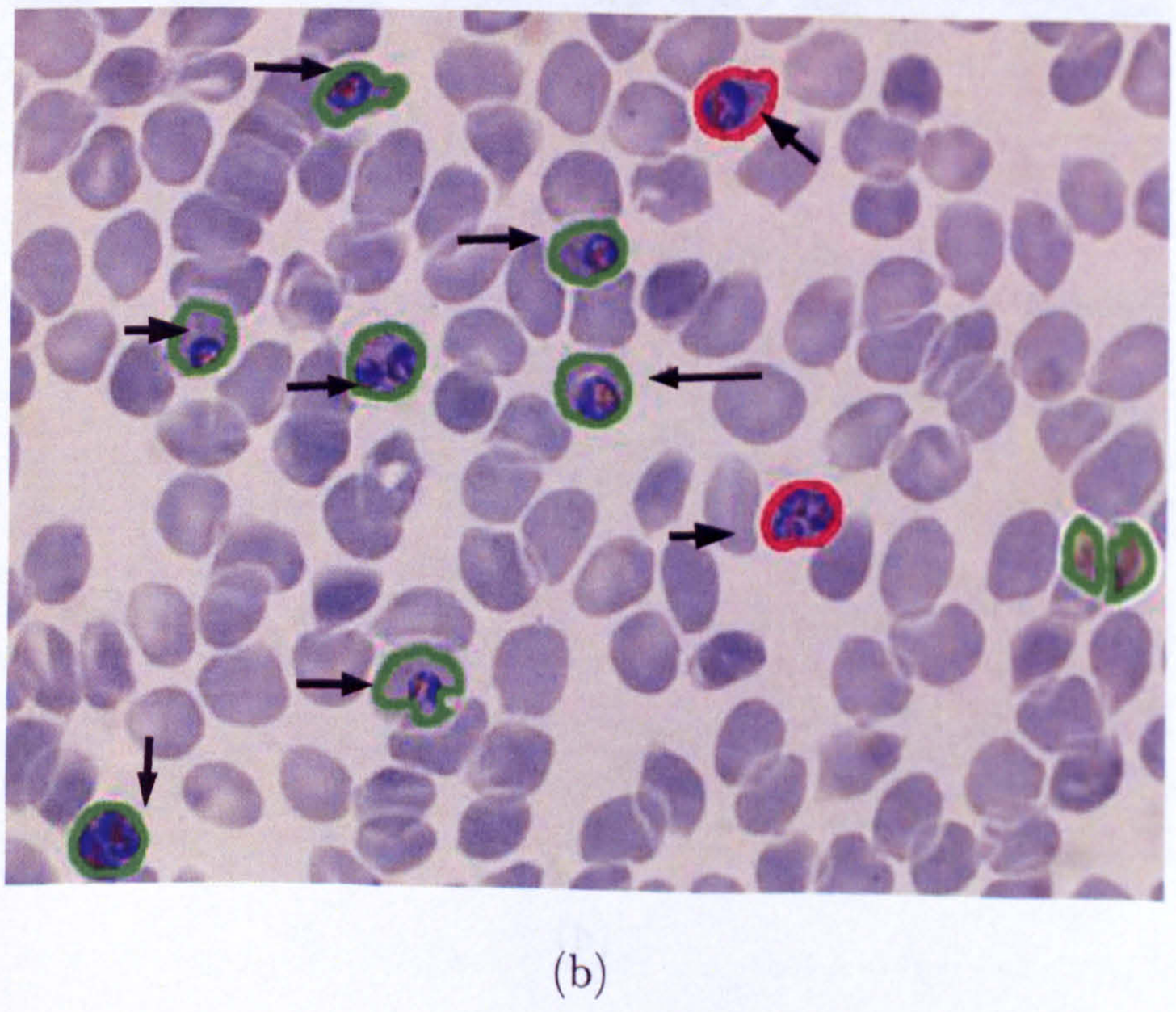
The scale dependent features (histogram and area granulometry of HGS) were able to cope with the arbitrary scale images in the test set Tt. Possibly, because there were objects at a similar scale in the training set Ta. However, as Figures 8.8 and Figure 8.9 show they are not as successful in images out of the scope of the sample sets. If we recall the training set experiments, we tried automatically scaling the images using the average cell area estimates (calculated by the area granulometry) of the images. Later, we discarded this procedure because the results were negatively affected. The degraded performance was possibly related to the inaccuracy of the average cell area estimation by the area granulometry. Additionally, it is possible that the objects at different scales were more easily distinguished by the scale sensitive features. Since we do not hold the real scale (i.e. pixel resolution) of the images, these hypotheses could not be tested.

However, it is possible to test the average cell area based scaling procedure in the test set Tt. Prior to the feature extraction every image in the image set was scaled with respect to its average cell area to construct the scaled feature sets: hgs^{Ta}, hgs^{Tt}, $sffs^{Ta}$, and $sffs^{Tt}$. Then these two sets were used together with the unscaled sets (HGS^{Ta}, HGS^{Tt}, $SFFS^{Ta}$, and $SFFS^{Tt}$) to observe the scale dependency.

Table 8.10 compares the performances of features extracted from the pre-scaled and un-scaled images. The same effects can be observed in both HGS and $SFFS$ features. As in the training set experiments the pre-scaling degraded the performance (compare *HGS-HGS* and *hgs-hgs* accuracies or $SFFS$-$SFFS$ and $sffs$-$sffs$). The pre-scaling procedure reduced the sensitivity of HGS by 4.2% (3.2% for $SFFS$). However, it can be observed that the unscaled training sets performed worse when the scaled test sets were used (see the *HGS-hgs* and $SFFS$-$sffs$ *Accs*). On the other hand, the scaled training sets performed more robust when the unscaled test sets were used (e.g.

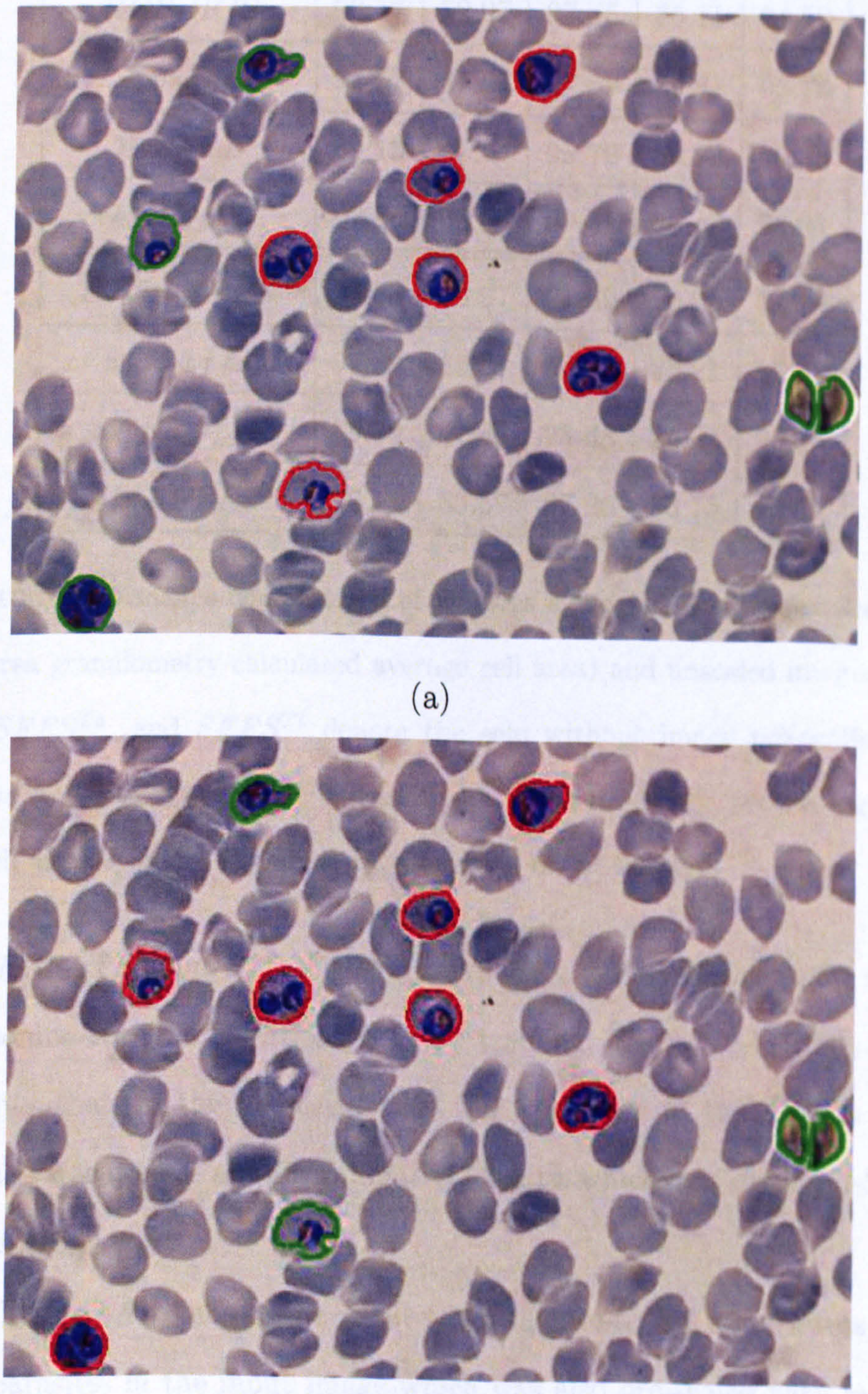


(a)



(b)

Figure 8.8: Parasite detection example images from other sources: (a) P. Vivax [10], (b) P. Falciparum (NIMR, London, UK). Arrows indicate parasites.



(a)

(b)

Figure 8.9: Parasite detection scale dependency: The detection results when the input image in Figure 8.8(b) is scaled by a factor of (a) 1.7, (b) with respect to its average cell area estimate.

Training^{Ta}-Test^{Tt}	Acc	SE	SP	PPV	NPV
HGS-HGS	93.17	72.07	97.45	85.11	94.52
hgs-hgs	92.67	67.87	97.69	85.61	93.76
HGS-hgs	89.13	41.74	98.72	86.88	89.33
hgs-HGS	92.06	63.66	97.81	85.48	93.01
SFFS-SFFS	93.02	69.97	97.69	85.98	94.14
sffs-sffs	92.32	65.77	97.69	85.21	93.38
SFFS-sffs	89.89	47.45	98.48	86.34	90.25
sffs-SFFS	92.11	66.67	97.26	83.15	93.51

Table 8.10: Performance comparisons of features extracted from pre-scaled (with respect to area granulometry calculated average cell area) and unscaled images. $IIGS^{Ta}$, HGS^{Tt}, $SFFS^{Ta}$, and $SFFS^{Tt}$ denote the sets without image pre-scaling whereas hgs^{Ta}, hgs^{Tt}, $sffs^{Ta}$, and $sffs^{Tt}$ denote the sets with image pre-scaling using the average cell area estimates (by area granulometry) of the images.

compare *SFFS-sffs* and *sffs-SFFS*). We have observed the same results with different training-test configurations, e.g. Tt training, Ta test set. This strengthens the possibility that, in the un-scaled case, the condition of the objects being from different scales was used as an (discriminative) utility, which aids distinguishing objects of different scales.

Finally, Figure 8.9(b) shows the result when the pre-scaled feature sets are used to detect the parasites in the input image which was also pre-scaled with respect to its average cell area estimate. It can be observed that more parasites were detected.

Although the pre-scaling procedure did not improve the performance, the result supports the necessity of a scale-invariant system. A system can still be trained and used in the same imaging conditions despite using scale-variant features. However, it will not be possible to exchange the images between the systems or more importantly

detected parasite signatures (i.e. features). Therefore, the scale information (i.e. pixel resolution in physical measures) should be recorded and provided by the imaging systems to enable accurate image scaling. The area granulometry based scaling can be used where this information is not available.

8.4 Discussions and Conclusions

- Sensitivity and Specificity

  Although we could not test the per-specimen (i.e. per-smear) accuracy of the proposed system, it is possible to provide a hypothethical comparison to manual microscopy. The average sensitivity threshold for manual microscopy (by an expert microscopist) of thin blood film examination is reported as 50 parasites/μl of blood [17]. According to Warhurst et al. [170] the sensitivity threshold of thin blood film examination is 11 times higher. Let us assume the average sensitivity threshold for manual microscopy (by an expert microscopist) of thin blood film examination as 500 parasites/μl. This corresponds to 0.01% based on the fact that average blood sample contains 5 million RBCs in $1\mu l$. This sensitivity figure is based on the assumption that expert microscopist works with 100% per-object sensitivity, however, can examine only a limited number of fields in a limited time. For example, if the microscopist examines 50 fields with average 200 RBCs, he/she would be able to examine 10000 RBCs, which should statistically have at least 1 infected cell for a specimen that is infected with at least 500 parasites/μl. According to the sensitivity result 72.1% of our parasite detector, its sensitivity threshold can be calculated as, i.e. approximately $\frac{1}{0.721} * 500 \approx 690$. Alternatively, to achieve an equivalent performance to the expert microscopy, the proposed detector must examine approximately 69 fields rather than 50. However, it should be noted that the routine positive/negative diagnosis is not performed on thin blood films. A study of routine malaria diagnosis in the UK showed that the

average detection sensitivity for microscopy was around 500 parasites/μl [24]. This may correspond to 5000 parasites/μl in thin blood films. Therefore, the detectors per-sample sensitivity suggests a performance comparable to the expert's and probably better than the average microscopist's. In addition, although we assume that the *expert* microscopist works with 100% per-object sensitivity, a recent study shows that the agreement rate among even expert microscopist's is not 100% and affected by the parasitemia level [171].

On the other hand, we have to consider specificity of the detector. According to medical statistics, $1\mu l$ of normal adult blood can contain 150,000 to 450,000 platelets and 4,500-10,000 white blood cells [69]. According to our sample set distribution, the number of artefacts can be higher than the platelets (as shown in Figure 9.2). Let us roughly assume that a healthy adult specimen of $1\mu l$ is prepared which contained about 1 million non-parasites (platelets + white blood cells + artefacts). If only 100 fields are examined, as in the sensitivity example, approximately 5000 thousand of these objects will be queried to the detector. The proposed detector with a specificity of 97.5 can be expected to produce a false positive result for every healthy specimen because it will incorrectly assign 1250 usual blood components or the artefacts as parasites. Hence, unless the specificity is increased, our detector would produce a trivial result: positive.

It is possible to increase the specificity by using the biased classifier. Thus, by reducing the sensitivity to ~23%, the detector can work with a specificity as high as 99.9% (see the ROC in Figure 8.3). However, this will raise the detector's sensitivity threshold to approximately 430 parasites/μl, which is not desired. This trade-off may be relived if some of these false positives can be eliminated by a post-interpretation. For example, based on the calculated distances a confidence rating scheme may be implemented. In addition, it must be noted that increasing number of parasite samples in the training set may improve.

Nevertheless, these figures are hypothetical. The method has to be tested and tuned on a significant number of both infected and healthy slides prior to deployment as a diagnosis tool. It must also be noted that a different application (not diagnosis) may require a different operating character. In this case, the variable sensitivity-specificity can be adjusted by the operator with the help of detector's ROC curve.

- The effect of the unbalanced training set

  Naturally, parasite detection is an unbalanced problem because there are more healthy stained components (i.e. WBC, platelet) than the infected RBCs. In addition, it can be expected that there would be more healthy people than infected. It can be expected that the cost of diagnosing an infected person/sample as healthy can be more than diagnosing a healthy person/sample as infected. However, although the costs can be proportional, it is not certain that the cost of a misclassification for the parasite class is more than the cost of a false detection. To diagnose a sample, our parasite detector has to consider hundreds of stained objects, not just a single one. Hence, the diagnosis result can be produced with a percentage or confidence scheme per slide. In order to reveal the diagnosis performance per slide, more experiments using different slides are necessary. By using the biased KNN classifier, which can adjust the sensitivity-specificity trade-off of the classifier, the system can be further tuned to produce desirable results for per-slide malaria infection diagnosis.

- Scale-variance

  The current evaluations reduced the scale invariance of the system by using scale-variant features. The scale normalisation based on the average cell area estimates (by area granulometry of the images) did not improve the results. However, it was shown that the procedure improves the robustness against scale variations. This was confirmed with some example results that were not included in our

sample sets.

The area granulometry based average cell area estimate may be inaccurate or in some conditions irrelevant as discussed in Sections 7.2.6 and 2.2. For example, in images of highly dense fields, RBCs considerably overlap. This makes it difficult to resolve an individual cell and estimate an average cell area. The alternative (and possibly the best) solution is to record and normalise the scale of the images with respect to the capturing imaging systems. The area granulometry based scaling should be used where this information is not available.

- Feature selection

  The sequential floating forward feature selection strategy effectively reduced the feature size without a significant trade-off of accuracy. When a query is made, the KNN classifier calculates distances to all training samples. Hence, as the number of training samples grows the speed of the classification can be a concern. In that case, the reduced set of features can be used.

- Sample Independence

  In Section 7.7 we defined four degrees of sample independence for the CDM application. The parasite detection experiments used the samples of the second degree, separating the samples of the *different images* into training and test sets. The *different slides* (i.e. third degree) could not be tested due to a limited number of slides. It must be noted the malarial parasite, *Plasmodium*, is divided into 16 sub-categories which noticeably differ with their morphological characteristics. It would not be possible, for example, to train with P. Falciparum rings, and then test with P. Vivax schizonts. However, the different slides were prepared and captured at different times. This provided a mix of different slides and imaging conditions (scale, colour properties) in the sample sets.

A KNN classifier was used for the classifications. The choice was motivated by

intuitiveness and some practical properties of the KNN classifier. KNN provides a good basis for future expansion of the diagnosis system. Furthermore, it can provide easy update schemes by simply adding new classes to the system if the features are relevant for the new parasite. A broader system, which can diagnose more than one parasite, will certainly be more useful than a system specific to only one parasite.

For the parasite detection, the most important function of the CDM, a variable detection accuracy scheme is proposed. In the per-slide-diagnosis scenario the specificity of the parasite detector can be more important than its sensitivity. The variable accuracy scheme operates on a sensitivity-specificity trade-off. Therefore, it enables a tuning for the per-slide-diagnosis scenario of the proposed detector, which currently produces the decision per-object. Hence, an important basis is provided for the realisation of the CDM.

The work presented in this chapter was presented and published by the author in [36,37]. A comprehensive publication, including the results shown here, is currently under preparation, to be submitted to the Journal of Medical Image Analysis (Elsevier).

Chapter 9

Diagnosis - Species and Life-Stage Recognition

Species recognition is the second important task in CDM because the medical treatment can differ from one species to the other. In case of a detected infection, usually further analysis is essential to determine the species. Determination of the life-stage(s) is also required because it can indicate the degree of the sickness. Additionally, the life-stage recognition can be useful for scientific studies where parasite development with respect to a treatment must be observed.

In Chapter 1, a general overview of the species and life-stage recognition problem has been given. For a microscopy diagnosis expert, the tasks of parasite detection, life-stage, and species recognition are not necessarily sequential or independent. A microscopy diagnosis expert can perform all these tasks in a single classification or sequentially or even partially depending on the discriminative information that exists in the observed object. For example, the expert can recognise a *P. Falciparum* ring stage parasite directly; or recognise a ring stage parasite and can seek for more discriminative parasites to decide the species. Moreover, it is not required to determine the life-stage of every single parasite because a thorough examination of the whole slide can reveal the most frequent life-stage and the condition of a single/mixed species infection. Therefore, if one considers the diagnosis of a whole slide, the species and life-stage recognition tasks can be regarded as contextual.

In this study, due to the limited number of thin blood film slides available, the contextual properties of the problem were not investigated. As for parasite detection, life-stage and species recognition tasks were considered to be based on a single stained object. In other words, the study proposed here does not jointly consider several objects in the same image or slide. This is identified as future work.

The malaria parasites are growing life forms. They gain more characteristics of the species as they mature. Hence, they can be more easily distinguished as they mature. Moreover, as explained in the Section 1.3.1, some particular life-stages parasites can be definitive for the recognition of species. For example, P. Falciparum gametocytes usually occur in a crescent shape, which is not seen in the other species. On the other hand, a ring stage (immature) parasite of all four species can be relatively similar. Hence, the life-stage and species recognition tasks should be investigated as separate but also as joint tasks.

In the previous detection discussions, the subclasses of the parasite classes were not considered; all the parasite species and life-stages were treated as a single class (i.e. parasite). The same was true for the non-parasite objects. Figure 7.1 demonstrated the sub-classes of the problem, which is subject to this discussion.

In Section 1.3.1, the most common properties of the different species were listed in Table 1.3.1 and the illustrations of different life-stages were shown in Figures 1.1-1.4.

A similar methodology to that for the parasite detection problem is used here. The detailed explanations about the methodology can be found in Chapter 7. The same data set is used here; however, the data is not divided into two sets. Section 9.1.1 presents some preliminary information about the sample distribution, which guided our experimental methodology.

9.1 Experiments

It is possible to perform several different experiments considering the hierarchy of the classes (Figure 7.1).

1. 20-Class All Samples Classification: This overrides the parasite detection task and it investigates whether the three tasks (i.e. detection-species-life-stage) can be performed by a single classification.

2. 16-Class Parasite Samples Classification: This assumes the parasite detection is

performed already. Therefore, it investigates whether the two tasks (i.e. species/-life-stage) can be performed jointly by a single classification.

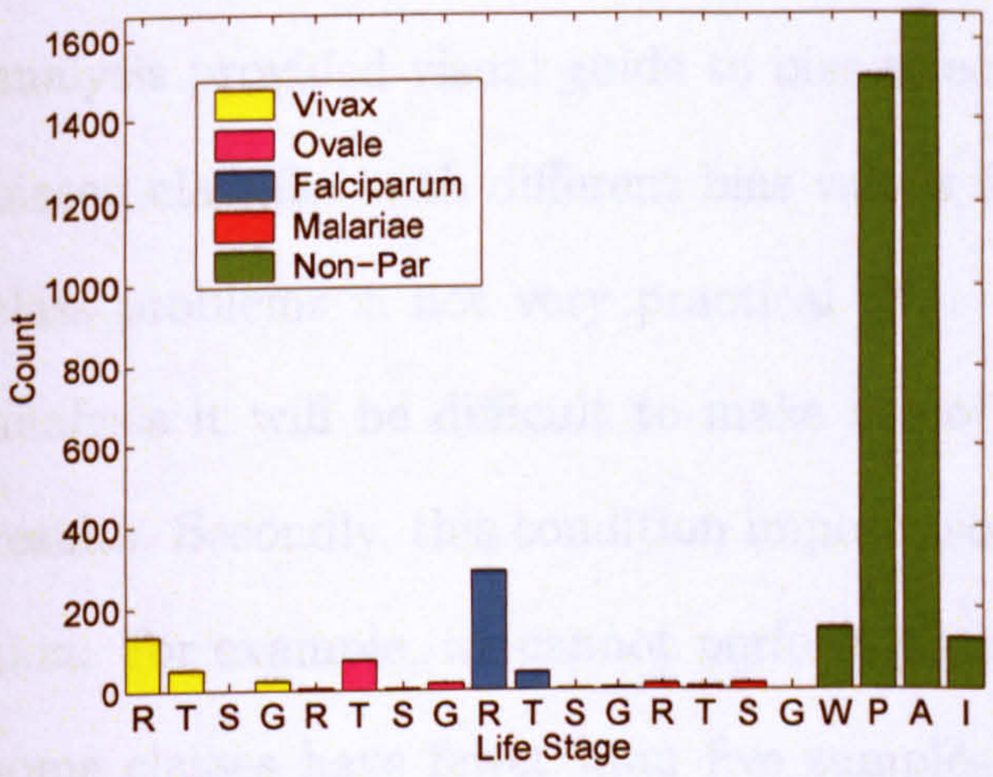
3. 4-Class Species and 4-Class Life-Stage Samples Classification: In these cases, all three tasks are treated separately. Hence, two classifications are performed separately for the species and life-stage categorisations.

The following section analyses the sample distribution for the classes and discusses the appropriateness of the different experiments.

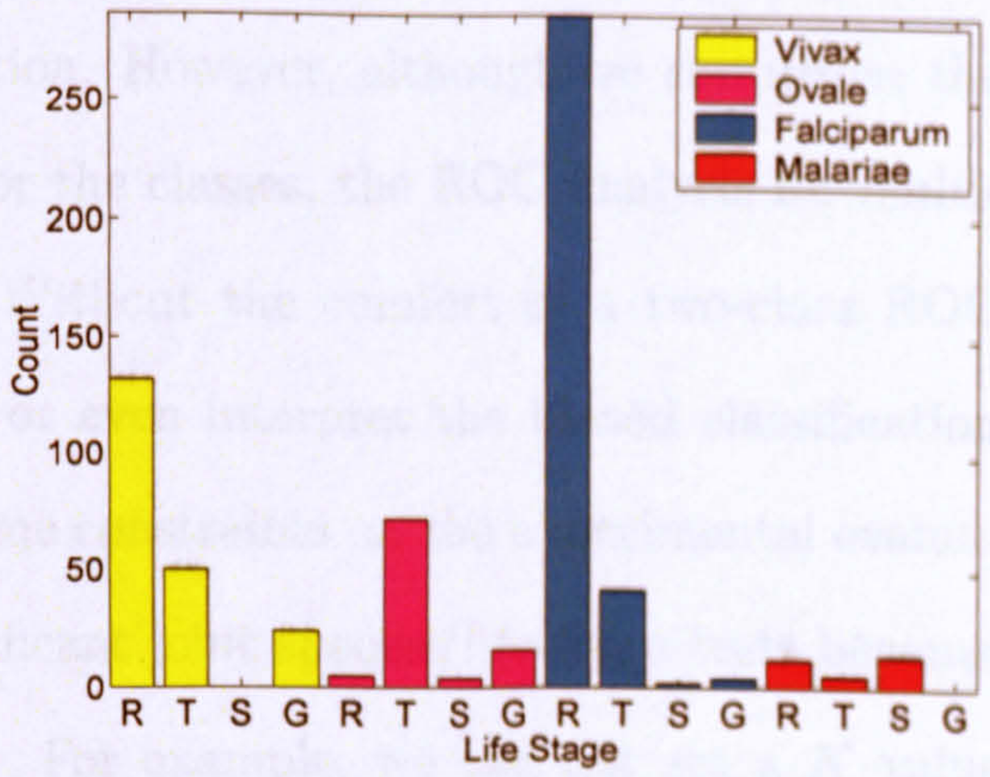
9.1.1 Preliminary

The information about the UoW thin blood film slides and existing parasite species was provided in Section 3.2. It will be appropriate to look at the species and life-stage distributions in the sample set. Here, all of the existing samples (i.e. Ta+Tt, see Section 8.1) were used. The histogram in Figure 9.1(a) presents the distribution of the samples into the subclasses (see also Figure 7.1). As we were aware from the detection problem, the non-parasites were the most populated class in the sample set. However, here, the distribution of the parasites and non-parasites can be seen in the lowest level (except white blood cell types). The most populated non-parasite class was the artefacts followed by the platelets, white blood cells and incomplete extracted objects. Figure 9.1(b) shows only the distribution of the parasite subclasses. The most populated parasite-species-(life-stage) class was P. Falciparum-rings followed by P. Vivax-rings, P. Ovale-trophozoites. It can be observed that the sample set did not contain any P. Vivax-schizonts and P. Malariae-gametocytes. The distribution of the samples into species and life-stage classes can be seen in Figures 9.1(c) and 9.1(d), respectively. The species distribution was as follows P. Falciparum(335), P. Vivax(206), P. Ovale(97), P. Malariae(31). The life-stage distribution was as follows Rings(436), Trophozoites(169), Gametocytes(44), Schizonts(20).

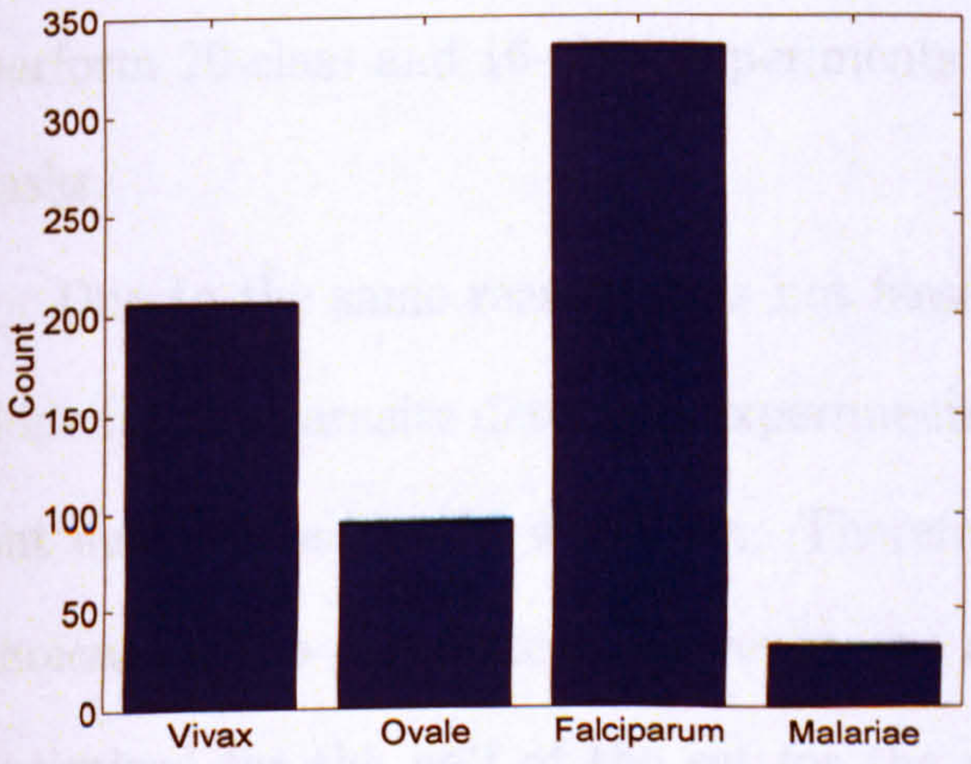
The unbalanced distribution of the samples and the lack of two life-stages of P.



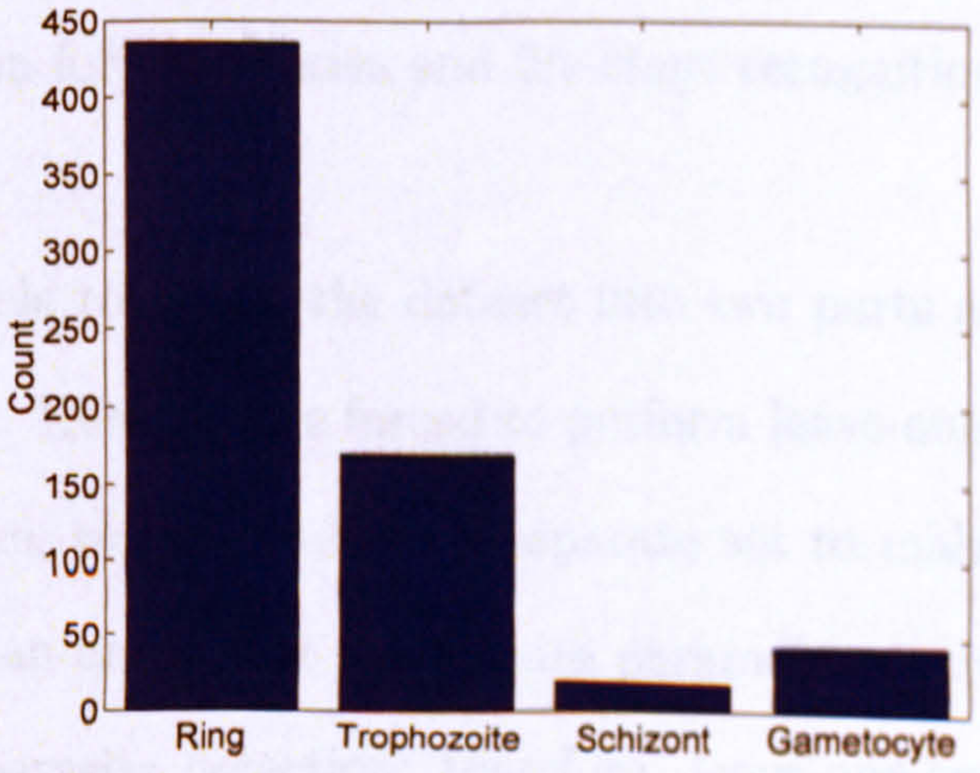
(a)



(b)



(c)



(d)

Figure 9.1: Distribution of the samples: (a) All data; sub-labels denote (R: ring, T: trophozoite, S: schizont, G: gametocyte, W: white blood cell, P: platelet, A: artefact, I: incomplete) samples. The histogram is rearranged to show only parasites (b); sum of species (c); sum of life-stages (d).

Vivax and Malariae is not very satisfying, particularly for the 20-class and 16-class experiments. The unbalanced distribution will cause the less frequent classes to be dominated by the others. This was faced also in the parasite detection problem; however, it was partially solved by using the biased KNN classifier for which the ROC analysis provided visual guide to bias selection. However, although we can utilise the biased classifier with different bias values for the classes, the ROC analysis for multi-class problems is not very practical [48]. Without the comfort of a two-class ROC analysis it will be difficult to make use of or even interpret the biased classification results. Secondly, this condition imposes some constraints on the experimental evaluation. For example, we cannot perform significant joint species/life-stage tests because some classes have fewer than five samples. For example, we can not set a K value of the KNN classifier to 5. However, it is still possible to ignore this condition and perform 20-class and 16-class experiments in for the species and life-stage recognition tasks.

Due to the same reasons it is not feasible to divide the dataset into two parts as done for the parasite detection experiments. Here, we are forced to perform leave-one-out evaluations on the whole set. Therefore, we do not have a separate set to make choices for the parameters. However, one can argue that the feature parameters were optimised for the half of the set for the parasite detection; therefore, leave-one-out evaluations can be optimistic. On the other hand, one can argue that the classification problems are different; so the use of the same features will be inefficient because features should be once again determined for the separate classifications. In fact the second argument is stronger: switching between a two-class and a multi-class problem is not only about altering the class labels. Provided that the features are the same, the samples will be located in the same positions in the Euclidean feature space as the two-class problem. However, the classifier decision boundaries are expected to relocate for the KNN and most classifiers. However, KNN has a special condition: when $K = 1$ (i.e. nearest neighbour) multi-class or binary classifications are equivalent. Because

when the feature space does not change the nearest neighbour of the query does not change.

Therefore, the best approach can be to re-evaluate the feature performances for the current task. However, none of the feature parameter tuning, feature comparison procedures were applied here. They were left for future studies where more samples available. Experiments were performed on the concatenated *IIGS* feature using a basic KNN classifier with the RL_1 distance metric.

A final issue, which is specific to KNN, is the *tie* (equal "votes") membership conditions for a decision. In the case of a binary classification, tie conditions are prevented by using odd K values: for an odd K value one of the two classes always has more nearest neighbours than the other. However, this is not possible for multi-class KNN: inevitably tie conditions occur when the training samples distributed to the different 20 classes are considered for a finite set of K neighbour. For a finite sample size the tie conditions are expected to increase with the number of different classes; however, they can be reduced in number by using large K values. However, large K values were not feasible for our sample set because there are few examples of some classes. Instead, the tie conditions were solved by assigning the queries to the nearest neighbour's class.

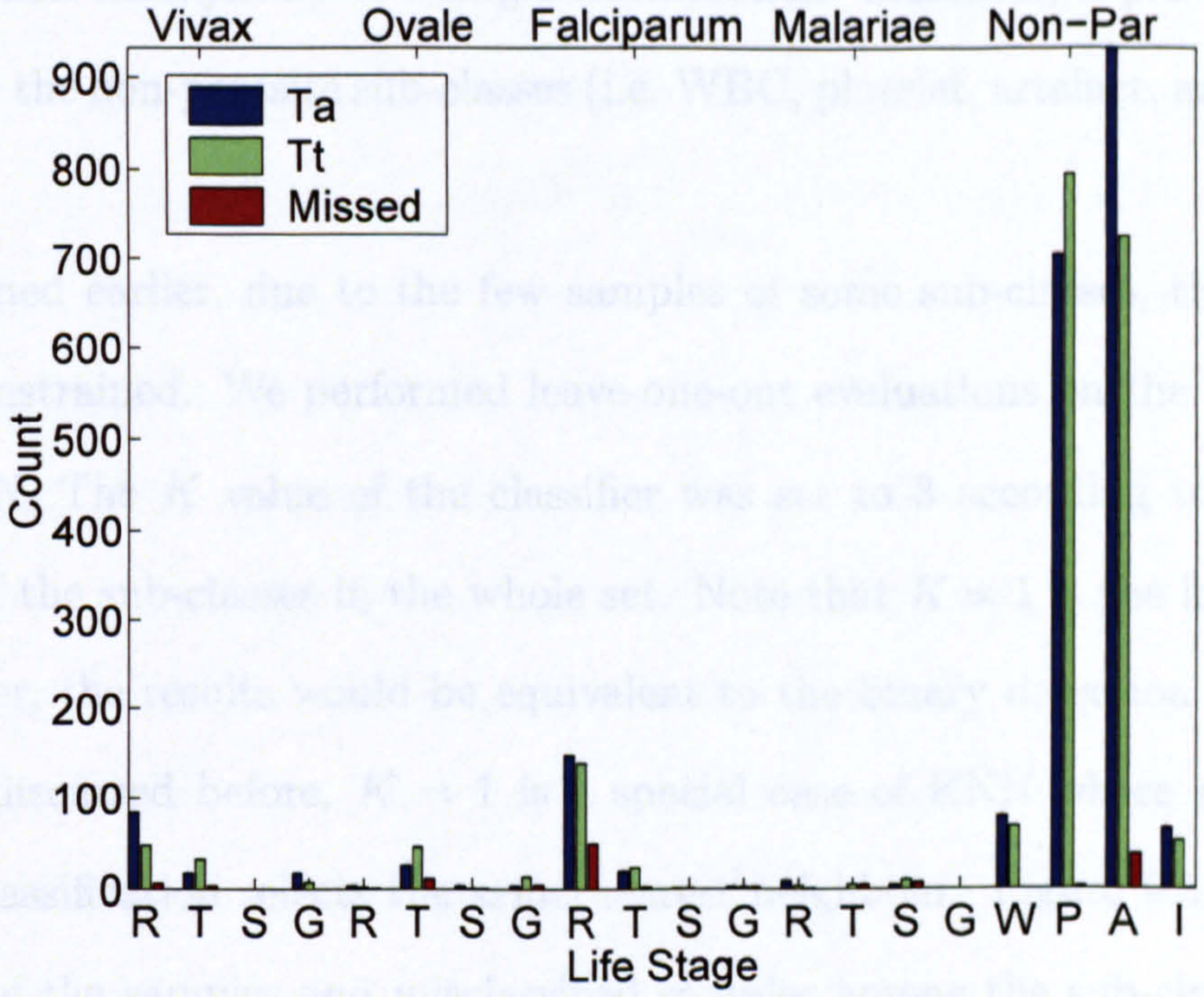
9.1.2 Parasite Detection Misclassifications

Before beginning the primary experiments, we can perform one more analysis which can contribute to the analysis of the misclassifications of the parasite detection task. Additionally, we can gain some additional insight in to the overall problem. The binary parasite detection experiments ignored the subclasses of the problem and treated all the species and life-stages as the parasite class whereas all the non-parasite samples were treated as the non-parasite class. Furthermore, the whole sample set was divided into two partitions then used for the hold-out evaluations. The partitioning ensured the

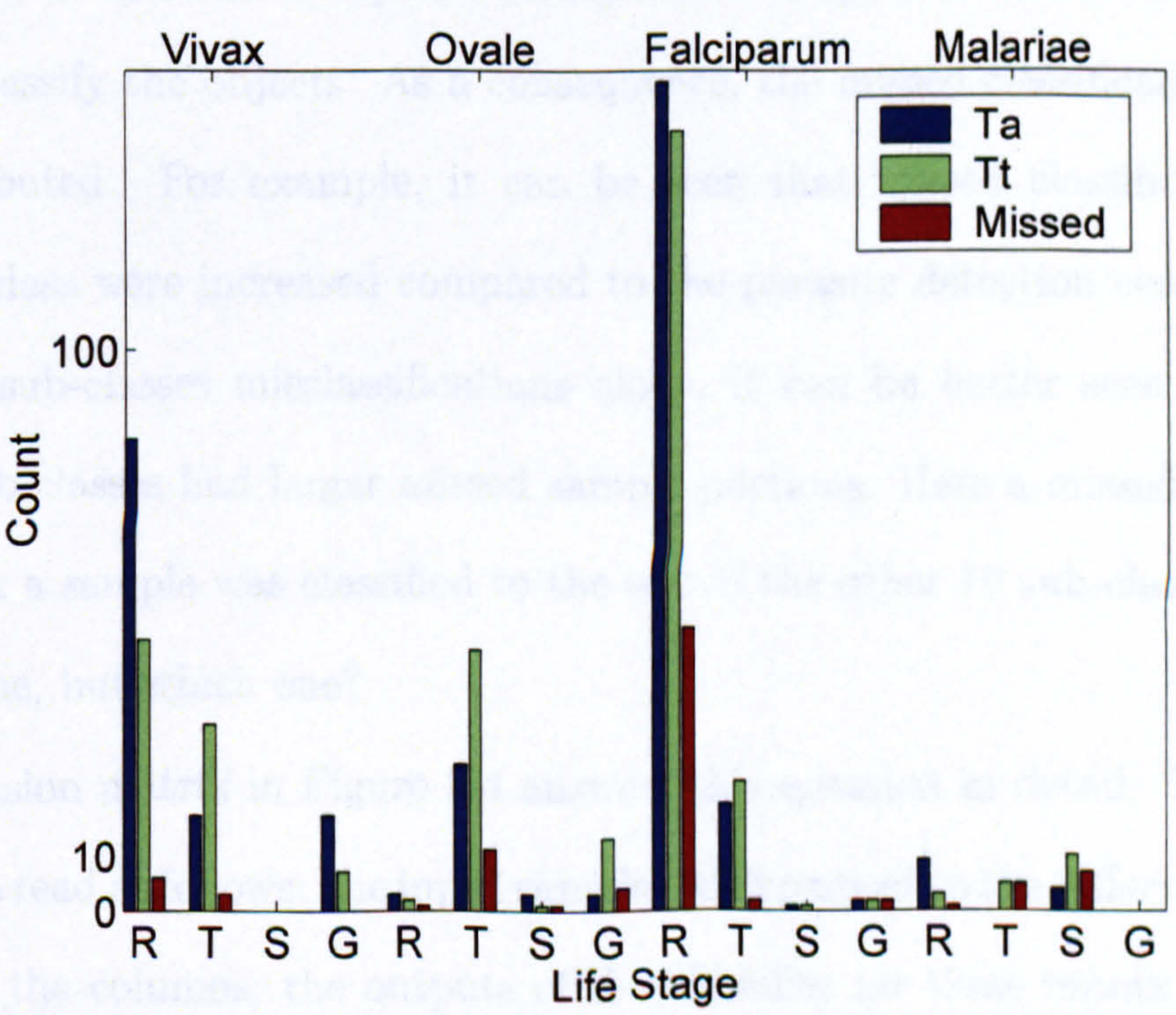
training (Ta) and the test (Tt) samples were from different images. However, there was no criterion to ensure that the sub-class samples were equally distributed among two sets. Figure 9.2(a) shows the distribution of the classes in the training (Ta) and test (Tt) sets, respectively. It can be seen that the samples (of sub-classes) were not equally distributed between the two sets. Moreover, in some of the sub-classes the test set samples were more than the number of the training set samples. This may be critical for the sub-classes which have fewer samples. For example, it can be seen that the training set did not contain any Malariae trophozoites (M.T). This may be the reason that all of the test samples from this sub-class were misclassified (Figure 9.2(a)). The classifier was designed to learn a generalised parasite notion without considering the sub-categories by capturing the similarities of the parasite objects and thus discriminating against the non-parasite objects. Because there were few samples of Malariae trophozoites, we can not conclude directly that all the Malariae trophozoites were classified as non-parasites due to lack of generalisation. There might be other factors which could cause this, e.g. the unbalanced parasite/non-parasite sample distribution. However, it is a possibility and an important issue to investigate further. If a generalisation problem exists, this may indicate that the features were not able to capture the variation of the objects successfully to represent the "*parasiteness*"/"*non-parasiteness*" of the objects, or the features were sufficient but the classifier could not develop a generalisation of the *parasiteness/non-parasiteness* notion. In either case, this may suggest that the generalisation should be sought at a different level rather than the highest parasite/-non-parasite level.

9.1.3 20- and 16-Class Classifications

The 20-Class classification regards every sub-class of the problem as independent. In contrast to the parasite/non-parasite binary classification, it seeks a generalisation in the lowest level in the hierarchy of the objects (according to our chart Figure 7.1). In



(a)



(b)

Figure 9.2: The parasite detection samples and misclassified samples' distributions among the subclasses: (a) All sub-classes, (b) Only the parasite sub-classes.

other words, it intends to perform the parasite detection, species recognition and life-stage recognition tasks jointly in a single classification. Moreover, it provides discrimination within the non-parasite sub-classes (i.e. WBC, platelet, artefact, and incomplete objects).

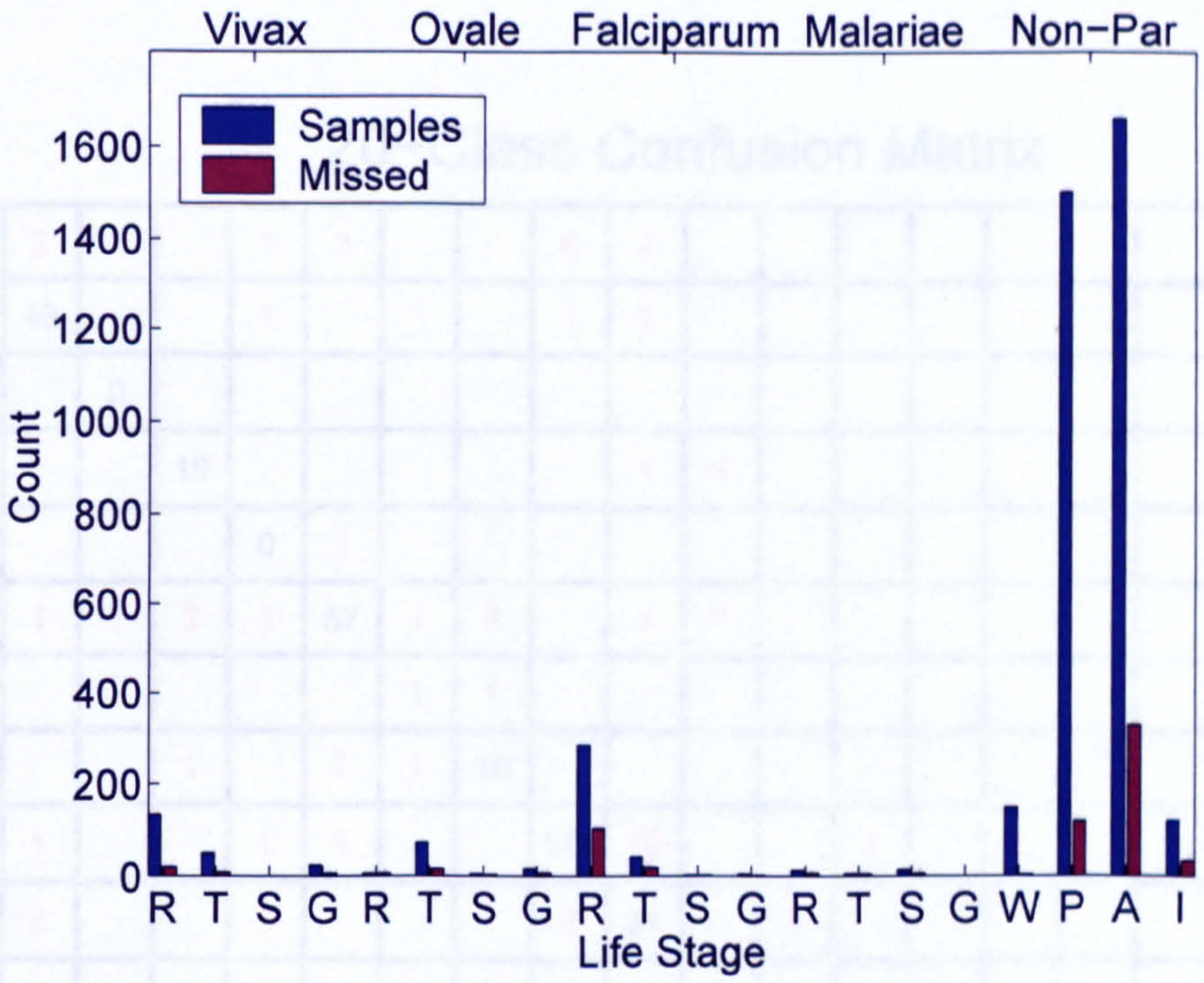
As explained earlier, due to the few samples of some sub-classes, the experiment had to be constrained. We performed leave-one-out evaluations on the whole sample set with KNN. The K value of the classifier was set to 3 according to the smallest population of the sub-classes in the whole set. Note that $K = 1$ is the lowest possible value, however, the results would be equivalent to the binary detection classification. Because, as discussed before, $K = 1$ is a special case of KNN where a two-class or multi-class classification selects the same nearest neighbour. Figure 9.3(a) shows the distribution of the samples and misclassified samples among the sub-classes. It must be noted that, in this case every sub-class had to compete with 19 other sub-classes in order to classify the objects. As a consequence, the missed classifications are more widely distributed. For example, it can be seen that missed classifications in the platelet sub-class were increased compared to the parasite detection case. Examining the parasite sub-classes misclassifications alone, it can be better seen that the less populated sub-classes had larger missed sample portions. Here a missed classification indicates that a sample was classified to the one of the other 19 sub-classes instead of the correct one, but which one?

The confusion matrix in Figure 9.4 answers this question in detail. The confusion matrix can be read as follows: the input samples with respect to the different sub-classes are shown in the columns; the outputs of the classifier for these inputs are shown in the rows; the diagonal terms show the correct classifications; the rest of the terms show with which other sub-class the patterns were confused. It is easy to read it in column order. For example, examining P. Vivax rings (the first column, i.e. V.R), we can see that 112 of the patterns were correctly classified whereas 1,10,2,2, and 5 of the P. Vivax rings were incorrectly assigned to the P. Vivax-trophozoite (v.T), P. Falciparum-ring,

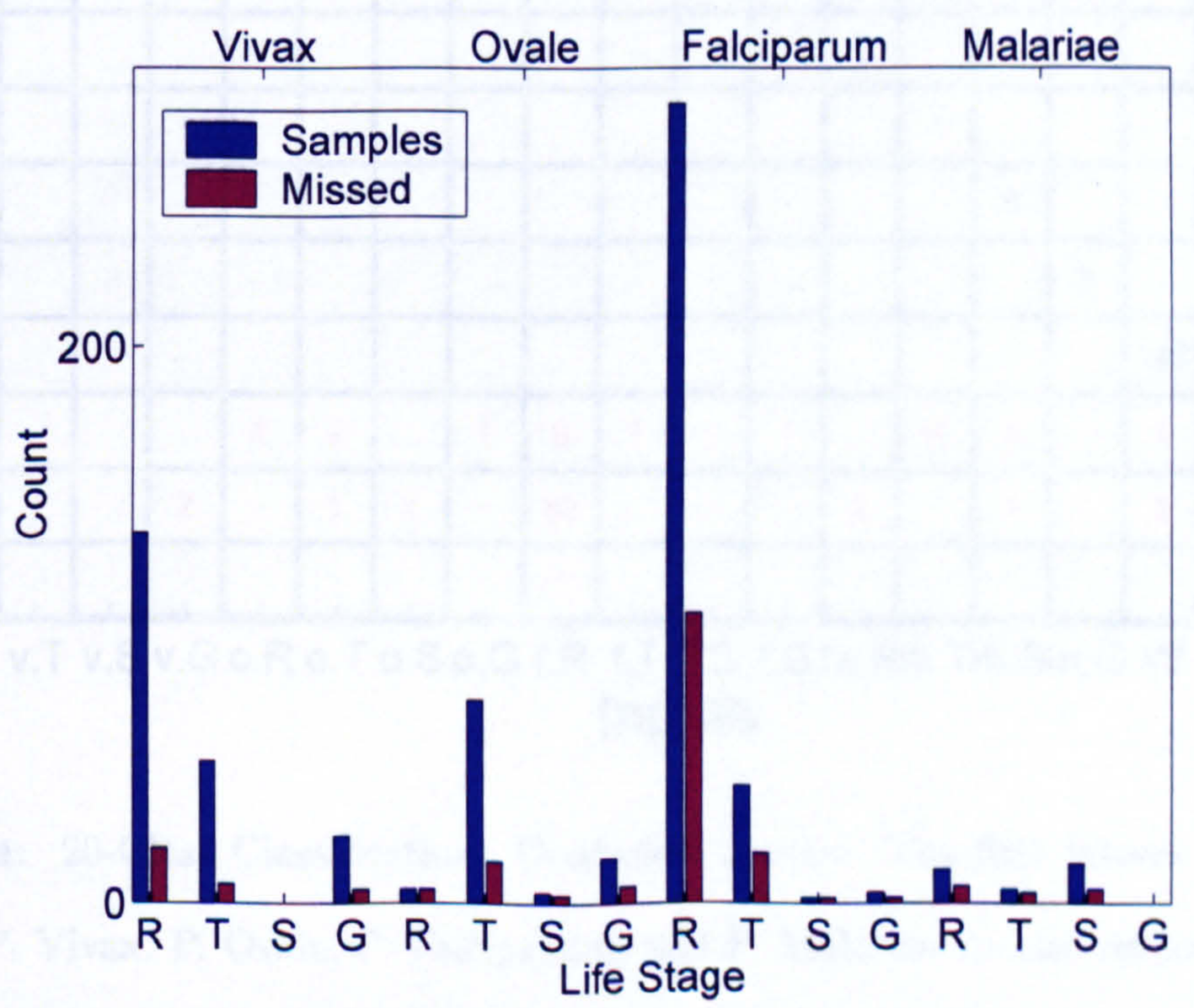
P. Falciparum-trophozoite, platelet, and artefact sub-classes respectively. There is a lot of detail in the confusion matrix; however, the sample distribution was unfair to treat all confusions as significant. Some remarkable observations can be summarised. Most of the parasite sub-classes were confused with the artefact sub-class of the non-parasites and vice-versa. The ring life-stages were the most confused sub-classes. Within the non-parasite sub-classes, most of the confusions occurred between the platelets and artefacts; within the parasite sub-classes most of the confusions occurred across the different species rings (e.g. V.R-F.R); and the life-stage confusions (e.g. F.R-F.T) within the same species were observable. Finally, it is remarkable that within the non-parasite sub-classes, confusions were limited by the artefacts and platelets and was not spread to the WBCs, which may enable the counting of WBCs.

In the ideal case, we should report the classification accuracy for each sub-class separately. If we consider that some of the sub-classes had few or no samples, the results would be far from significant. Additionally, the lowest level accuracies were not a major concern here. However, we can still calculate the binary parasite detection, species and life-stage recognition accuracies by categorising the errors for the sub-classes. These were calculated and compared in the Section 9.1.6.

The 16-class classification assumes that a binary parasite/non-parasite classification, the detection, is performed previously. It actually performs the same joint species/life-stage classification as the 20-class classification; however, the non-parasite samples are discarded beforehand by the binary detection process. All the classifier settings were the same as 20-class classification. Figure 9.5 shows the distribution of the samples and misclassified samples among the 16 parasite sub-classes. It can be observed the most missed sub-classes were P. Falciparum-trophozoites, P. Falciparum-rings, P. Vivax-rings, and P. Ovale-trophozoites, respectively in the order. The confusion matrix in Figure 9.6 reveals how these missed classifications were assigned. It can be observed that most of P. Falciparum-trophozoites were assigned to the P. Falciparum-rings class. There were other within-species confusions such as between P. Ovale-trophozoites and



(a)



(b)

Figure 9.3: 20-Class Classification: Distribution of the samples and missed classifications among the sub-classes, (a) All sub-classes, (b) Only the parasite sub-classes.

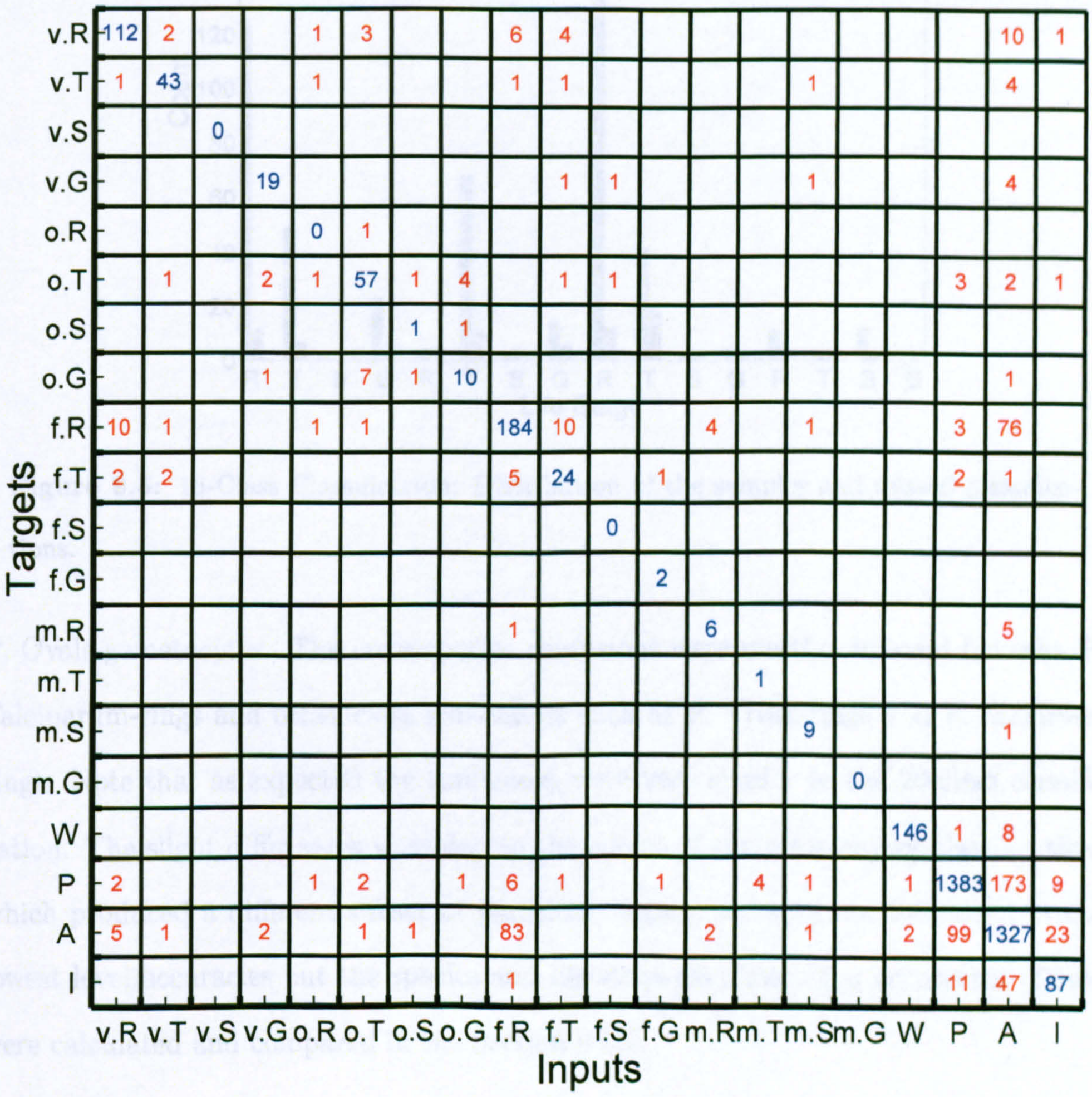


Figure 9.4: 20-Class Classification: Confusion matrix. The first letters v, o, f, and m denote P. Vivax, P. Ovale, P. Falciparum, and P. Malariae species respectively. The second letters R, T, S, and G denote Ring, Trophozoite, Schizont, and Gametocyte life-stages respectively. W, P, A, and I denote white blood cell, platelet, artefact and incomplete objects respectively.

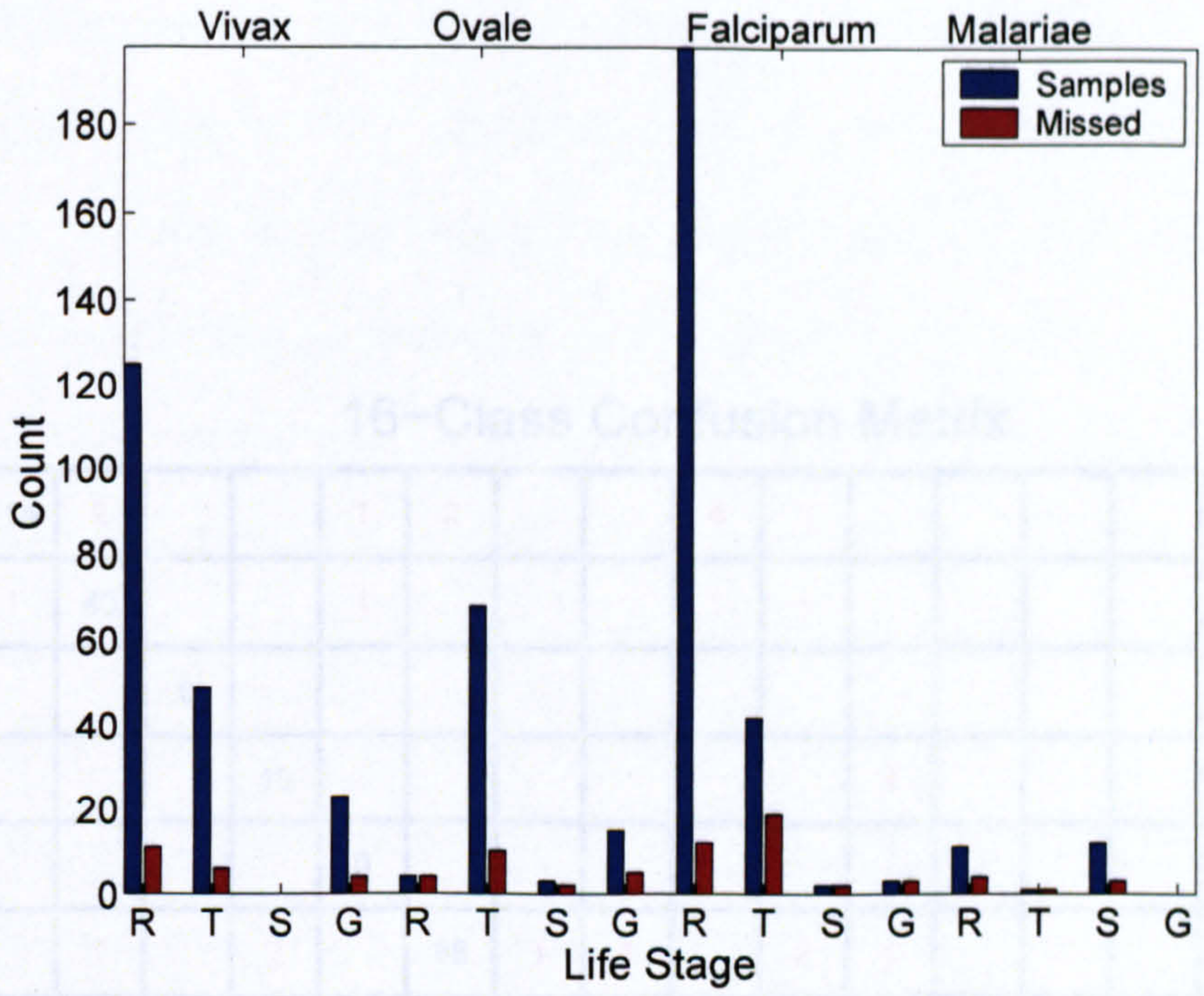


Figure 9.5: 16-Class Classification: Distribution of the samples and missed classifications.

P. Ovale-gametocytes. The cross-species confusions were mostly observed between P. Falciparum-rings and other close sub-classes such as P. Vivax-rings and P. Malariae-rings. Note that as expected the confusions were very similar to the 20-class classification. The slight differences were due to the effects of the prior binary classification, which produced a different subset of parasites. Again, we were not interested in the lowest level accuracies but the species and life-stage discrimination accuracies. These were calculated and compared in the Section 9.1.6.

9.1.4 4-Class Species Classification

The 4-Class species classification assumes that a binary parasite detection is performed beforehand. It seeks a generalisation in the species level. In other words, the classifier intends to develop a generalisation for the species classes (P. Vivax, P. Ovale, P. Falciparum, P. Malariae) and discriminate them. Hence, after the binary parasite detection, the original sample labels are reproduced by discarding the lowest life-stage levels and

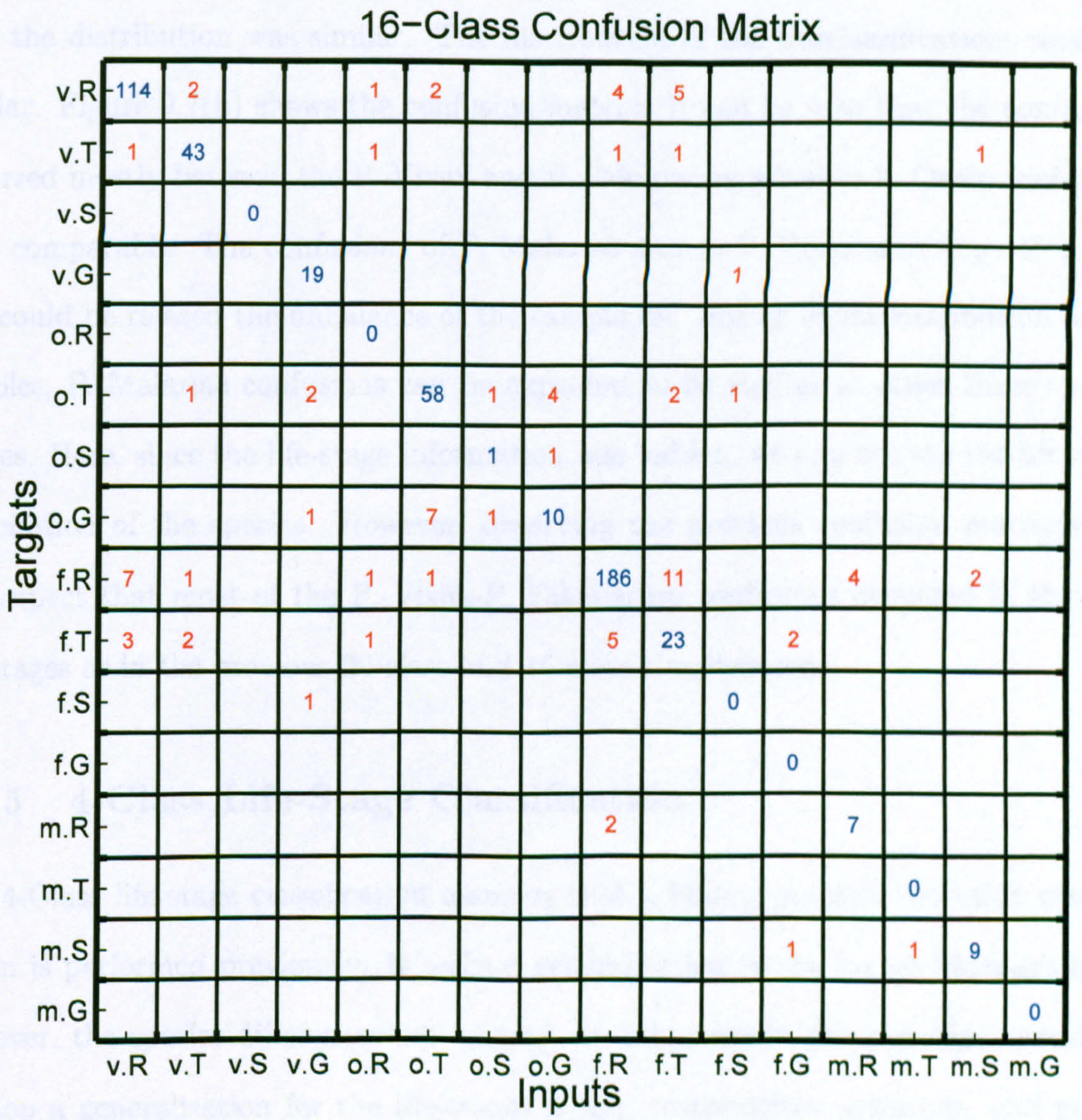
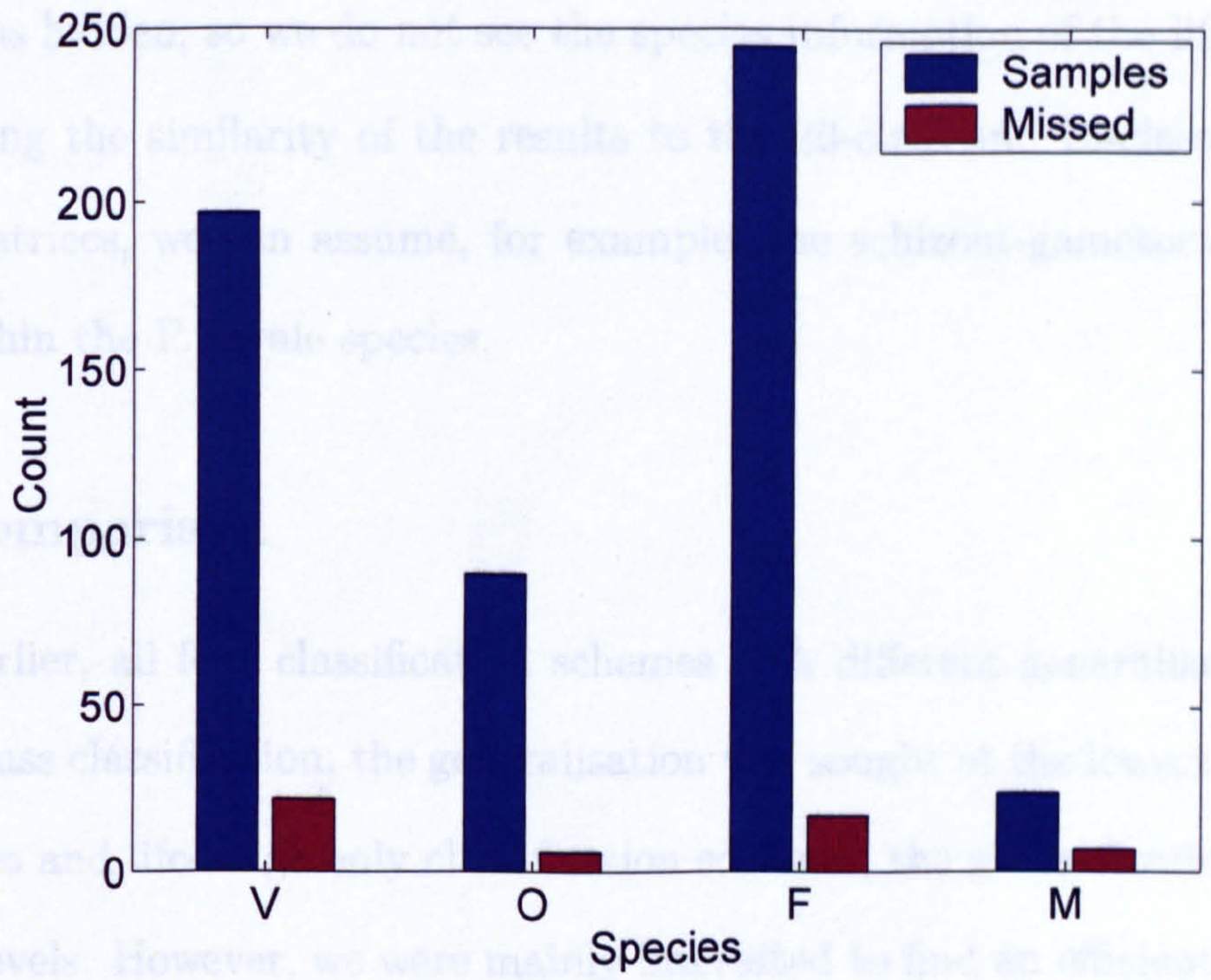


Figure 9.6: 16-Class Classification: Confusion matrix. The first letters v, o, f, and m denote P. Vivax, P. Ovale, P. Falciparum, and P.Malariae species respectively. The second letters r, t, s, g denote Ring, Trophozoite, Schizont, and Gametocyte life-stages respectively.

tested in the KNN classifier with the same settings as before. Figure 9.7(a) shows the sample and missed classifications' distributions for the set of detected samples by the binary parasite detector. The preliminary experiments showed previously the distribution of the samples into species classes in the whole sample set. Here, it can be observed that the distribution was similar. The distribution of the misclassifications was also similar. Figure 9.7(b) shows the confusion matrix. It can be seen that the confusions occurred mostly between the P. Vivax and P. Falciparum whereas P. Ovale confusions were comparable. The confusions of P. Malariae and to P. Vivax were low. However, this could be related the unbalance of the sample set. For an equal distribution of the samples, P. Malariae confusions can be expected to be similar to other three species classes. Here, since the life-stage information was hidden, we can not see the life-stage information of the species. However, observing the previous confusion matrices, we can expect that most of the P. Vivax-P. Falciparum confusions occurred in the ring life-stages as in the previous 20-class and 16-class classifications.

9.1.5 4-Class Life-Stage Classification

The 4-Class life-stage classification assumes that a binary parasite detection classification is performed previously. It seeks a generalisation in the lowest life-stage level. However, the species differences are ignored. In other words, the classifier intends to develop a generalisation for the life-stages (rings, trophozoites, schizonts, and gametocytes) by considering for example P. Vivax-ring and P. Falciparum-ring in a single class. Figure 9.8(a) shows the sample and missed classifications' distributions. Note that a binary parasite detection classification was applied prior to the classification. Again, the distribution of the samples was not equal. Figure 9.8(b) shows the confusion matrix. It can be seen that the confusions occurred mostly between the Trophozoite (T) and Ring (S) classes. The second most common confusion was the opposite case. It was plausible to observe that the confusions between morphologically distant classes



(a)

Species Confusion Matrix

Targets \ Inputs	V	O	F	M
V	175	3	11	1
O	8	84	4	
F	14	3	228	6
M			2	17

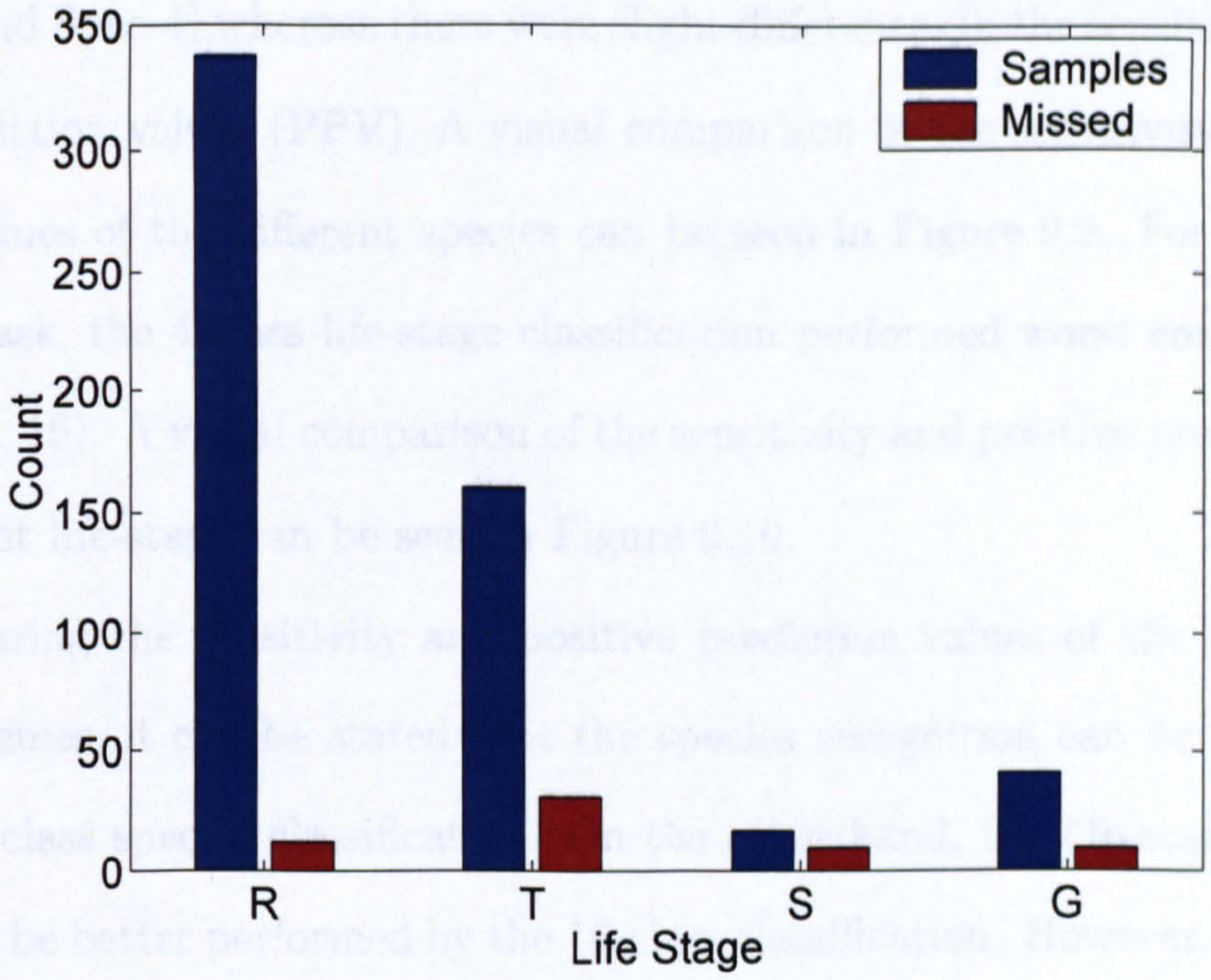
(b)

Figure 9.7: 4-Class Species Classification: (a) Distribution of the samples and missed classifications, (b) confusion matrix. The letters V, O, F, and M denote P. Vivax, P. Ovale, P. Falciparum, and P. Malariae species respectively.

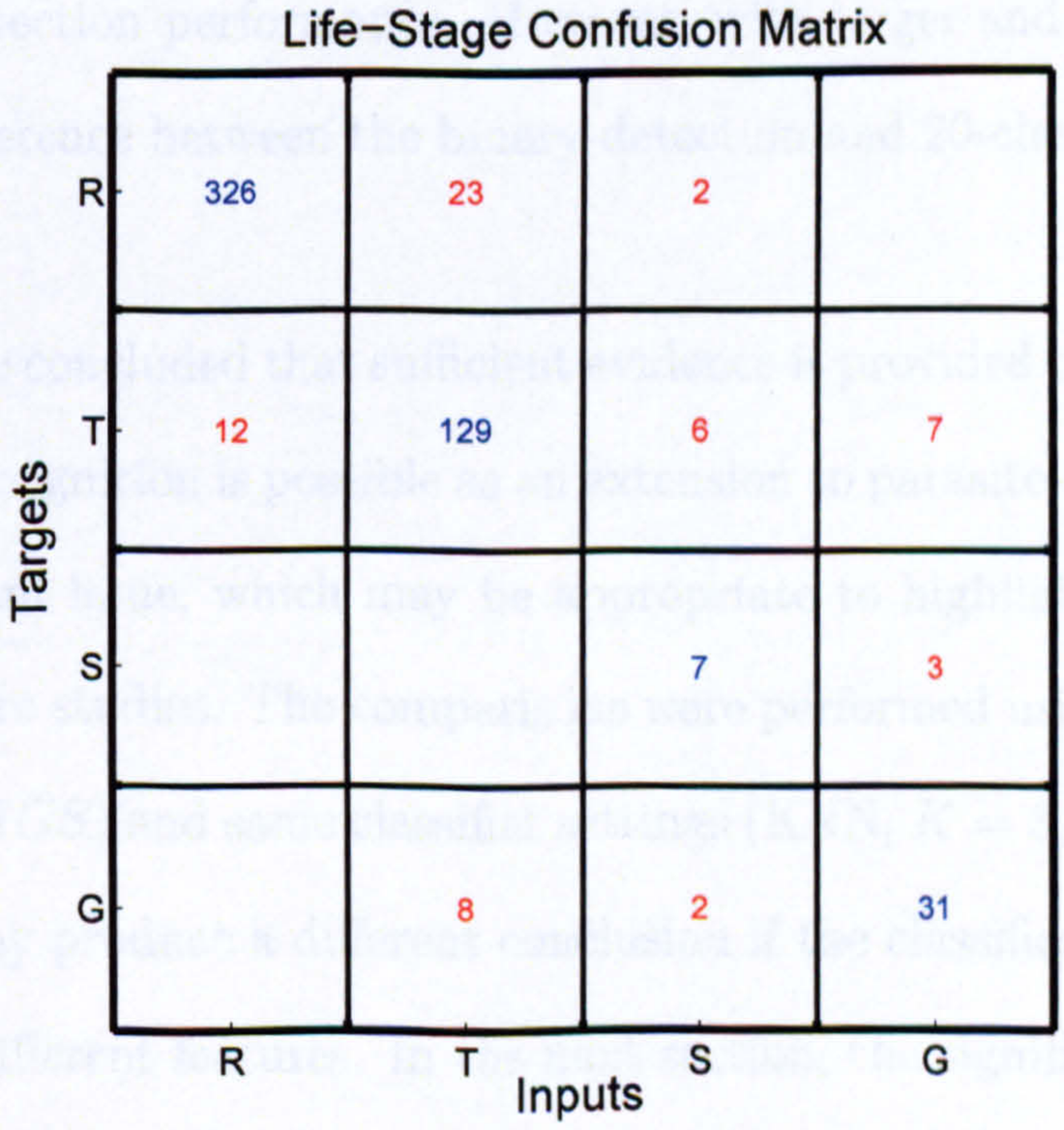
(i.e. schizonts(S)-rings(R) or gametocytes(G)-rings(R)) did not occur. The species information was hidden, so we do not see the species information of the life-stage. However, observing the similarity of the results to the 20-class and 16-class classification confusion matrices, we can assume, for example, the schizont-gametocyte confusions occurred within the P. Ovale species.

9.1.6 Comparison

As stated earlier, all four classification schemes seek different generalisations. In the 20- and 16-class classification, the generalisation was sought at the lowest levels. In the 4-class species and life-stage only classification schemes, the generalisation was sought in different levels. However, we were mainly interested to find an efficient classification scheme to perform only species and life-stage recognition. Thus, although the confusion matrices observed so far give enough evidence to see that there is no significant difference, we calculated the life-stage and species recognition performances (i.e. Acc, SE, and SP) and compared these numerically for the respective schemes. Table 9.1 summarises the numerical results. It must be noted all results were obtained by the leave-one-out evaluations on the whole set. Hence, the parasite detection performances are slightly higher compared to the hold-out ($Ta - Tt$) evaluations of the parasite detection experiments. It can be observed that the detection performances of the 20-class classification and the binary parasite detector were similar. However, this is related to the very low K setting (i.e. 3). The results would be equal in the $K = 1$ case because the nearest neighbour to a query does not change. We observed for the higher K values the 20-class parasite detection performance degrades. The number of detected parasites was close but the detected samples were different. Therefore, the 20-class classification scheme is not directly comparable to the other three schemes. However, its pre-detection performed equivalent, the 16-class classification, can be compared to the 4-class classifications. For both species recognition task, it can be observed that



(a)



(b)

Figure 9.8: 4-Class Life-Stage Classification: (a) Distribution of the samples and missed classifications, (b) confusion matrix. The letters R, T, S, and G denote ring, trophozoite, schizont, and gametocyte life-stages respectively.

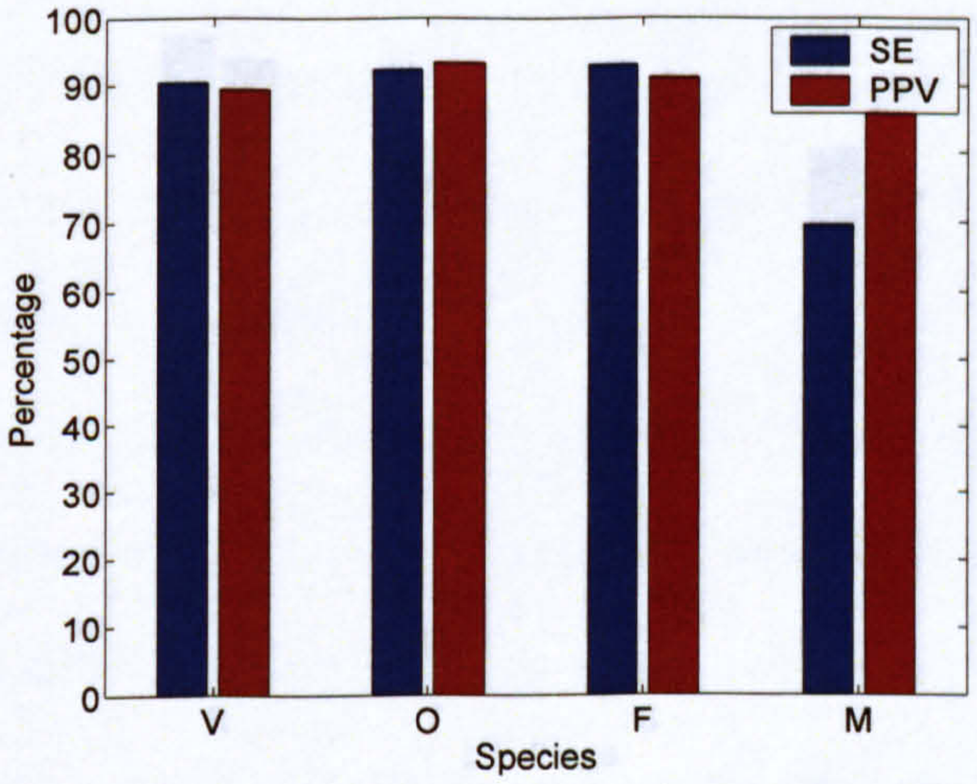
the average accuracy (Acc) values were not significantly different for the three schemes (i.e. 20, 16, and Spec-4) whereas there were slight differences in the sensitivity (SE) and positive prediction values (PPV). A visual comparison of the sensitivity and positive prediction values of the different species can be seen in Figure 9.9. For the life-stage recognition task, the 4-class life-stage classification performed worst compared to the other two (20, 16). A visual comparison of the sensitivity and positive prediction values of the different life-stage can be seen in Figure 9.10.

By comparing the sensitivity and positive prediction values of the different classification schemes, it can be stated that the species recognition can be performed by 16-class or 4 class species classification. On the other hand, the life-stage recognition task seems to be better performed by the 16-class classification. However, it is also possible to perform 20-class classification and perform all three tasks at once with a cost of slightly reduced detection performance. However, with larger and equally distributed sample sets the difference between the binary detection and 20-class classification may be reduced.

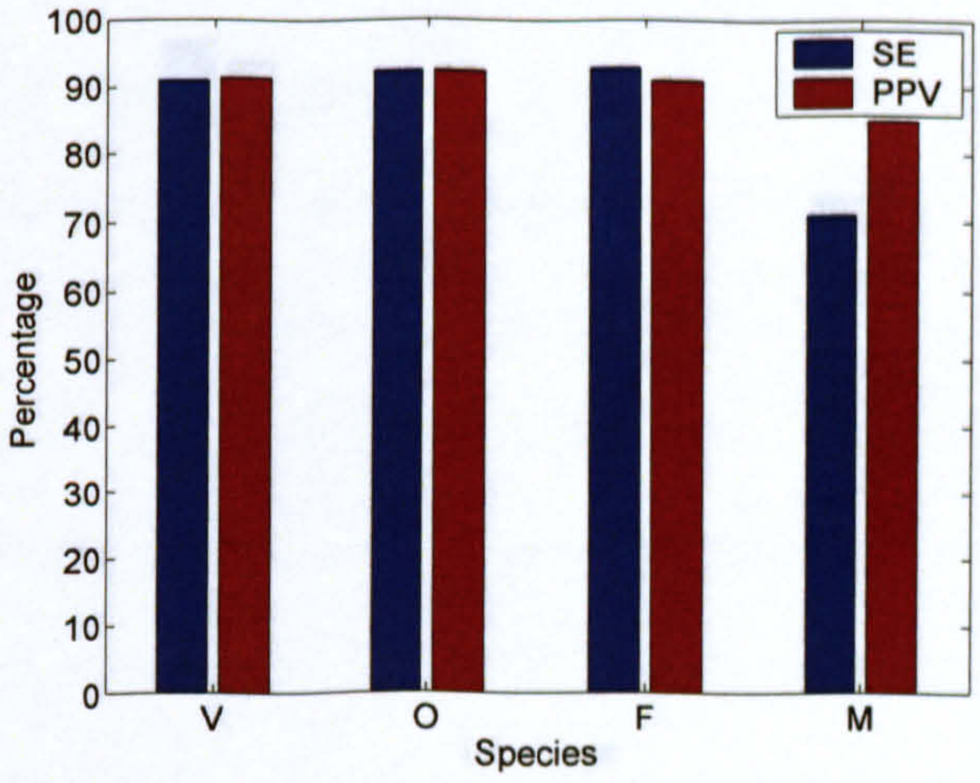
Finally, it can be concluded that sufficient evidence is provided to show that the life-stage and species recognition is possible as an extension to parasite detection. However, there is an important issue, which may be appropriate to highlight here and may be investigated in future studies. The comparisons were performed using the same sample set, same feature (HGS) and same classifier settings (KNN, $K = 3$, and RL_1 distance). The comparison may produce a different conclusion if the classification schemes could be allowed to use different features. In the next section, the significance of the results is discussed with a broader computerised diagnosis perspective.

		Binary	20	16	Spec-4	Life-4
Detection	Acc	94.5	94.4			
	SE	83.1	82.7			
	SP	96.8	96.7			
	#Detected Parasites	556	553	556	556	556
Species	Acc		91.0	91.0	90.7	
	P. Vivax SE		90.3	90.9	88.8	
	P. Vivax PPV		89.4	91.3	92.1	
	P. Ovale SE		92.3	92.2	93.3	
	P. Ovale PPV		93.3	92.2	87.5	
	P. Falciparum SE		93.0	92.7	93.1	
	P. Falciparum PPV		91.1	90.8	90.8	
	P. Malariae SE		69.6	70.8	70.8	
	P. Malariae PPV		94.1	85.0	89.5	
Life Stages	Acc		89.9	89.2		88.7
	Ring SE		96.7	96.5		96.5
	Ring PPV		93.4	93.1		92.9
	Trophozoite SE		81.3	81.3		80.6
	Trophozoite PPV		86.1	85.0		83.8
	Schizont SE		58.8	58.8		41.2
	Schizont PPV		90.9	71.4		70.0
	Gametocyte SE		80.0	73.2		75.6
	Gametocyte PPV		74.4	76.9		75.6

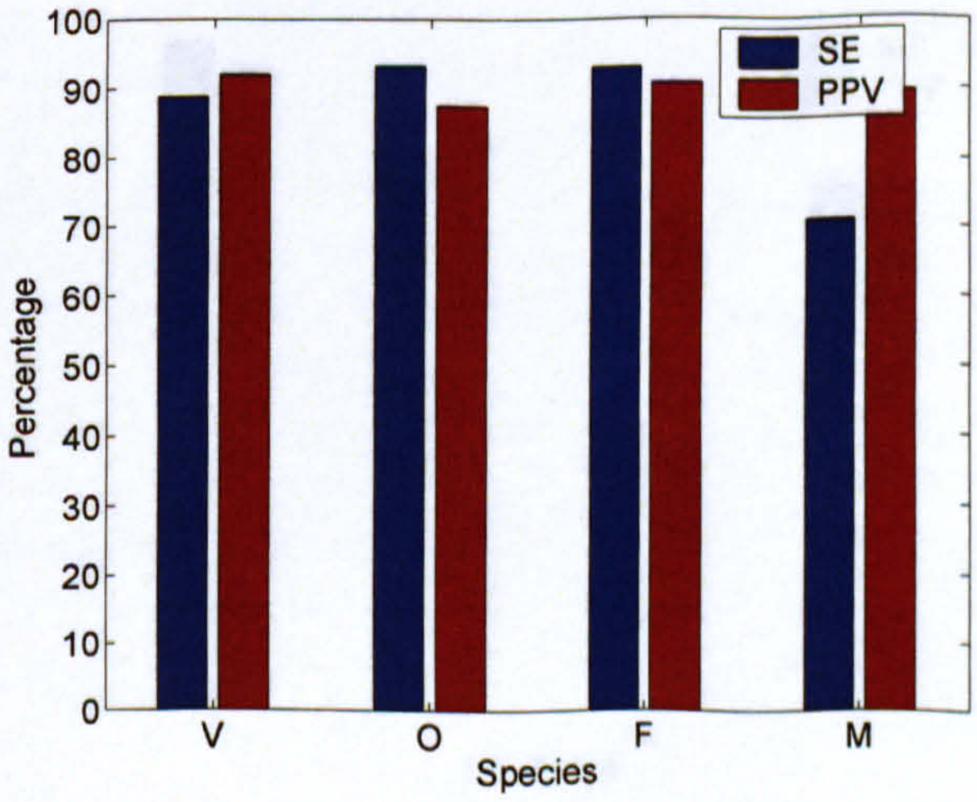
Table 9.1: Comparison of the 20-class, 16-class, and 4-class (Spec-4 and Life-4) classification schemes for the species and life-stage recognition tasks. SE: sensitivity, PPV: positive prediction value.



(a)

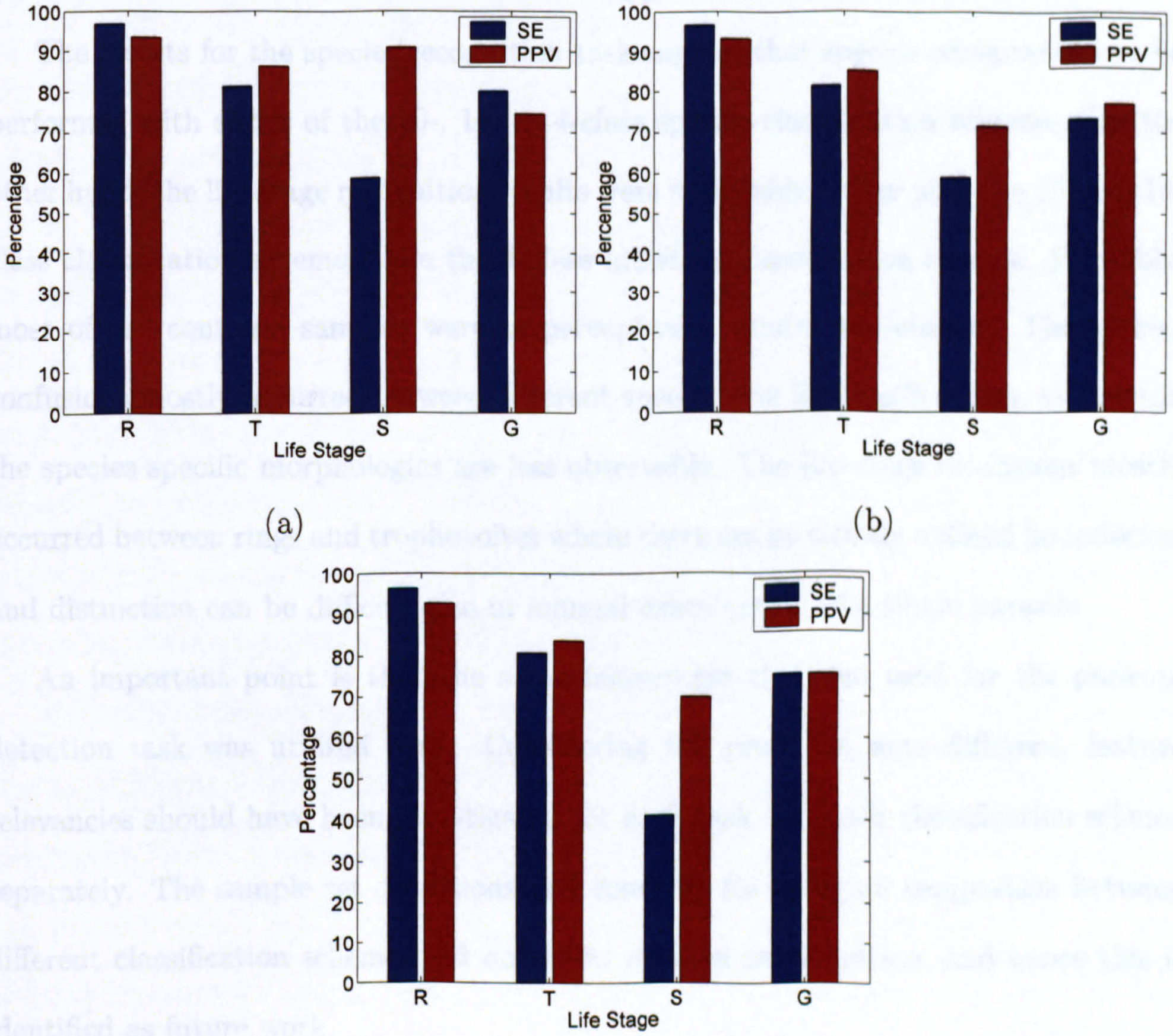


(b)



(c)

Figure 9.9: Comparison of sensitivity (SE) and positive prediction values (PPV) of the species recognition task for different classification schemes: (a) 20-class classification, (b) 16-Class classification, (c) 4-Class species classification.



(a)

(b)

(c)

Figure 9.10: Comparison of sensitivity (SE) and positive prediction values (PPV) of the life-stage recognition task for different classification schemes: (a) 20-class classification, (b) 16-Class classification, (c) 4-Class life-stage classification.

9.2 Discussions and Conclusions

The species and life-stage recognition tasks were seen as extensions to parasite detection. However, the study considered the three tasks as both consecutive and joint. The experimental results have shown that either approach can provide a solution.

The results for the species recognition task suggest that species recognition can be performed with either of the 20-, 16- or 4-class species classification schemes. On the other hand, the life-stage recognition results were noticeably better with the 20- and 16-class classification schemes than the 4-class life-stage classification scheme. Plausibly, most of the confused samples were in perceptually similar sub-classes. The species confusions mostly occurred between different species ring life-stages where, in general, the species specific morphologies are less observable. The life-stage confusions mostly occurred between rings and trophozoites where there are no strictly defined boundaries; and distinction can be difficult also in manual examination of a single parasite.

An important point is that the same feature set that was used for the parasite detection task was utilised here. Considering the problems were different, feature relevancies should have been investigated for each task and each classification scheme separately. The sample set conditions and concerns for an equal comparison between different classification schemes did not allow such an investigation, and hence this is identified as future work.

The unequal distribution of the species and life-stages in the sample set constrained the empirical evaluations for the species and life-stage recognition tasks. The comparison of the different classification schemes can be regarded as preliminary. However, they have shown promising evidence that species and life-stage recognition tasks can be performed as an extension to the parasite detection. It may also be possible to perform all three tasks in a single 20-class classification scheme.

In Section 7.7, we have defined four different degrees of sample independence to reveal thoroughly the generalisation for the CDM application. The experiments per-

formed leave-one-out evaluations on the whole sample set. Therefore, the training and test samples were allowed to be from the same images (i.e. the first degree of sample independence). The third degree of sample independence requires training and test samples to be from different thin blood film slides. This can reveal more generalised performance for all three tasks. However, in order to achieve this, the sample set must include a significant number of thin blood film slides. This has been identified as the primary future work.

In manual microscopy examination, the species and life-stage recognition tasks can be performed contextually for a whole slide. Therefore, the appearance of one parasite can be used as a clue for another parasite. Moreover, if a significant number of a single species is identified on a test slide, considerably less frequent species can be regarded as misclassifications. However, mixed infections should be taken into account. Unfortunately, due to the limited number of slides this issue could not be investigated. It can be investigated when a significant number of slides are available.

Although it was not one of the main objectives of the study, the 20-class classification experiments showed that also the non-parasite stained objects could be classified into the categories such as white blood cells, platelets, and artefacts. White blood cell and platelet counts of blood are important measures that are used in clinical analysis of many sicknesses. Despite being taken here as a whole category, the artefact class can include some blood anomalies (e.g. iron deficiency). The study can be extended to identify these conditions as well.

A comprehensive publication including the species and life-stage recognition studies is currently under preparation.

9.2.1 On the Related Studies

In two earlier studies performed by the University of Westminster group, [34, 35] the parasite detection task was not addressed. The analysis in [34] eliminated the WBCs

from the images and performed only life-stage analysis for P. Falciparum species. The analysis in [35,109] performed life-stage analysis directly on the stained objects without considering the presence of non-parasites. In comparison, the study presented here has taken a significant step towards CDM.

In a more recent publication [74] (April 2006, almost concurrently with our publications [36,37]) the problem was addressed as a consecutive (detection-species recognition) problem which is similar to approach taken in this study. The reported results were as follows (with our results of binary parasite detection and 16-class species and life-stage recognition shown in brackets): The results for parasite detection were reported as: Sensitivity (SE) of 85.1% (72.1%) with a Positive Prediction Value (PPV) of 80.8% (85.1%). The specificity value or false detection rate was not reported. See Section 7.5 for descriptions of these measures and note that PPV is a population dependent performance measure. For the species recognition task the SE-PPV results were reported as follows: *P. Falciparum* 57%-81% (92.7-90.8%), *P. Vivax* 64%-54% (90.9-91.3%), *P. Ovale* 85%-56% (92.2-92.2%), *P. Malariae* 29%-28% (70.8-85.0%). The life-stage recognition problem was not investigated.

Since the sample sets are different, a comparison of the two methods based only on the numerical results alone is unfair. The methods must be trained and tested on the same set of images for a fair comparison. However, simple numerical comparison would indicate that the methods presented here are comparable to, and possibly better than, Ross's.

However, we can methodologically compare the two approaches. The main differences are the pre-processing steps for obtaining stained objects, the choice of features, and the classifier algorithm. They followed Di Ruberto *et al.*ś approach [34] to extract stained objects (see Section 2.5). This was subsequently improved by Rao *et al.* [35] and further improved within this thesis study (see Section 6.6). In the classification stage, they proposed the use of two different sets of features for parasite detection and species recognition. They extracted 75, 117 individual features for parasite detection

and species recognition. Using principal component analysis (PCA [33]) they reduced the set of features to 37, 38, respectively. The use of optimised features may improve the performance but it degrades the method's expandability towards other parasites or blood anomalies. They trained two different two level BPNNs for the parasite detection and species recognition consecutively. We have shown in the detection experiments that BPNN has no advantage over KNN. Moreover, BPNN is not expandable, and it requires a new training (and possibly a new design) if a new function is desired (e.g. detection of a new blood parasite).

A more recent study [110] (Jan 2007) proposed "parasite counting" software. Similar to Rao's study [35, 109], they have experimented only on P. Falciparum *in vitro* images where the parasites are grown under a controlled environment. Other regular blood components and artefacts, which are the main problems of diagnosis, are not present in *in vitro* samples. Thus, the proposed program counted the stained objects and compared the results to manual counting. They stated that the program was intended for research purposes; however, not applicable to diagnosis. The species and life-stages recognition problems were not studied.

Chapter 10

Conclusions

This work reports the investigation of the diagnosis of malaria infection in peripheral thin blood film images. The problem was approached from the general computer vision, mathematical morphology, and statistical pattern recognition perspectives. In this chapter a summary, concluding remarks and future directions are provided.

10.1 Summary

The data collection methods used in this study were explained in Chapter 3. The implementation of the UoW microscope-camera setup was described. A set of 630 thin blood film images were acquired. The set contained examples of all four species of Plasmodium, which infect humans. The problem of non-uniform illumination was discussed. Two simple and effective solutions were implemented: the illumination image subtraction and morphological top-hats based filtering as proposed in [29]. In addition to image acquisition, the ground-truth data collection methods were explained. Three simple programs were implemented to aid manual markings. The collected data were then used in the segmentation, stained pixel detection, and object classification experiments.

The segmentation of the thin blood film images was examined in Chapter 4. The methods introduced a new modified watershed transform and a cell marker extraction technique based on the Radon transform. It also utilised existing tools (i.e. area granulometry [45], morphological double thresholding [35]). The proposed algorithm performed better than the earlier best algorithm [95], but did not overcome the common limitation: the segmentation algorithm was not applicable to images of dense fields. This study contributes to mathematical morphology by introducing the minimum area watershed transform. Its application to other increasing attributes such as

the area of a minimum enclosing rectangle, volume, or height can use the proposed algorithm by replacing the area criterion. Moreover, a generalised attribute watershed algorithm also covering non-increasing attributes (e.g. circularity, shape, and colour) can be possible. This potential should be investigated further.

The segmentation approach was not successful for the images of the dense fields of the blood film where blood cells do not have resolvable boundaries. Hence, the top-down deductive analysis strategy was altered to allow all images to be processed successfully. This was performed by searching and segmenting the objects locally in the images. The segmentation process was bypassed. This is illustrated in Figure 10.1.

Therefore, the diagnosis scheme was reduced to three main steps: stained object extraction, parasite detection, and species and life-stage recognition. The colour information in the RGB vector form was used extensively in all the steps. In order to use different images with different colour characteristics, colour constancy (i.e. colour normalisation) was studied.

Chapter 5 introduced a new colour normalisation method for thin blood film images. The method first separated the input image into foreground and background regions (i.e. image binarisation, [35]), then performed the normalisation sequentially. The method utilised the simple (and database) grey-world colour normalisation technique. However, it exploited simplicity of thin blood film images which are comprised of the plasma and cells (i.e. foreground objects). The effectiveness of the approach was demonstrated by the experiments. Moreover, examples demonstrated that the method could also work with images from unknown sources and images containing other blood parasites. Also further tests in the stained pixel detection experiments showed that the proposed colour normalisation method improves the Bayesian pixel classifier performance. It is possible to prepare a set of images (of the same field) under different acquisition conditions and compare the proposed method to the other general-purpose colour normalisation methods [129]. This is identified as future work.

Chapter 6 introduced a new local analysis method for detection and extraction of

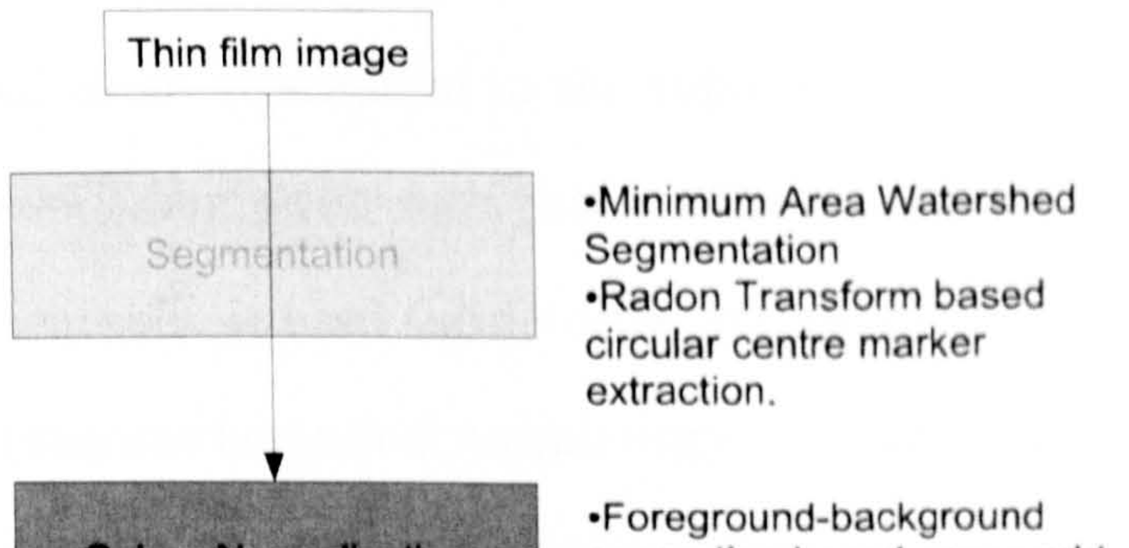


Figure 10.1: The proposed malaria diagnosis system. First, images are normalised for colour constancy. Then stained pixels are detected using a Bayesian classifier. Then they are processed locally using mathematical morphology to obtain stained objects. The detected stained objects (both binary masks and colour channels) are processed to extract the features. These are classified with K-Nearest Neighbour algorithm to discriminate non-parasites. The detected parasites are further processed to identify the species and life-stages. The global segmentation process is not part of the final method.

stained objects. The stained pixel detection task was approached as a two-class colour pixel classification problem. A Bayesian colour pixel classifier was implemented using the non-parametric histogram-based density estimation. The experiments showed that the method was successful in detecting stained pixels. The stained object extraction algorithm was developed as an alternative to the top-down segmentation strategy. It did not rely on a top-level assumption such that all the objects in the images are red blood cells and are segmentable. It used the detected stained pixels as the starting point and enlarged them by using mathematical morphology operations (i.e. area top-hat and reconstruction by dilation). By limiting the enlargement with respect to average cell area estimate (calculated by area granulometry [45]), the algorithm was able to extract the stained cell regions on the lightly populated fields, whereas in the densely populated fields the errors due to unresolvable boundaries were limited.

The proposed Bayesian pixel classification scheme made decisions based on single pixel colour and without considering local characteristics or other global input image conditions. Hence, it had to be supported by the local examination using morphological operations. However, it may be worth investigating whether stained pixel detection can be performed by a local analysis. For example, instead of using a single pixel colour vector to decide whether it is a stained or non-stained pixel, it may be possible to consider the neighbour pixels in a fixed size window. If stained pixel detection can be improved, it may be possible to bypass the morphological operations, which rely on the average cell area estimate. Therefore, the method would be also applicable to thick film analysis or other blood parasites. This is also identified as future work.

Chapter 7 explained the various aspects of the statistical pattern recognition methodology that was used in our diagnosis experiments: parasite detection, species and life-stage recognition.

Chapter 8 investigated parasite detection. It processed stained objects further to extract several colour- and shape-based features. It utilised a well-known supervised learning algorithm (K-Nearest Neighbour) as a classifier. A thorough experimental

study investigated some important aspects of the features and the classifier architecture. The results (Sensitivity: 72.1, Specificity: 97.5, Positive Prediction Value: 85.1, Negative Prediction Value: 94.5) show that parasite detection was possible using KNN (with the relative distance metric) with a concatenated feature comprised of colour histogram, area granulometry, and some simple shape measurements.

The scale-variance of the detection method was investigated. The average cell area estimate (calculated by area granulometry of the images) was used to perform image pre-scaling prior to feature extraction. Although the results were not higher, the comparative experiments suggested that the pre-scaling process improves robustness against scale changes. The area granulometry based average cell area estimate may be inaccurate or in some conditions irrelevant as discussed in Sections 7.2.6 and 2.2. For example, in images of highly dense fields, RBCs overlap considerably. This makes it difficult to resolve the individual cells and estimate an average cell area. The alternative (possibly the best) solution is to record and normalise the scale of the images with respect to the capturing imaging systems. However, the area granulometry based prescaling process should be used where this information is not available.

The evaluations of the detection method were based on a single stained object classification (i.e. per-object). A stained blood film slide contains hundreds of thousands of objects to be classified, and a real diagnosis scenario requires a per-smear (i.e. per-slide) decision. The per-object-classification results suggest that the per-smear-diagnosis can easily be performed. However, it may require a different sensitivity-specificity operating characteristic than the one used here. Hence, a variable sensitivity-specificity scheme, which can be used with the help of a ROC (Receiver Operating Characteristics) curve, was implemented. The variable scheme allows the classifier to adapt to the per-smear diagnosis scenario.

Parasite detection is the most important function for malaria diagnosis. The diagnosis is possible only if the system can differentiate between a parasite and a regular blood component. This thesis provided a parasite detection method, which can effec-

tively perform this task. However, a recent research study [74] addressed the problem with a similar approach to this study. Compared to our results, the reported detection sensitivity was higher, the positive prediction value was lower; however, the specificity value was not reported. We have shown that the sensitivity value alone is not sufficient to suggest acceptable diagnosis performance (see Section 8.4). Moreover, our biased KNN classifier allows a variable sensitivity-specificity trade-off. For example, if 73% percent specificity (SP) is acceptable, the classifier can provide 97% sensitivity. Therefore, a direct comparison based on the numerical results is not fair. The methodological differences between the methods were discussed in Section 9.2.1.

Chapter 9 investigated species and life-stage recognition. These tasks were seen as extensions to the parasite detection. The possibility of performing all three tasks in single joint classification was also investigated. The unequal distribution of the species and life-stages in the sample set constrained the empirical evaluations for the species and life-stage recognition tasks. However, the experiments have shown promising evidence that species and life-stage recognition tasks can be performed as an extension to the parasite detection. In addition, it was shown in this thesis that it is possible to perform all three tasks in a single joint classification.

However, the same feature set that was used for the parasite detection task was utilised for species and life-stage recognition. Considering the problems were different, feature relevancies should have been investigated for each task and each classification scheme separately. The sample set conditions and concerns for an equal comparison between different classification schemes did not allow such an investigation. This is identified as future work.

Moreover, although it was not one of the objectives of the study, it was shown in species experiments that non-parasite objects could also be classified into categories such as white blood cells, platelets, and artefacts. White blood cell and platelet counts of blood are important measures that are used in clinical analysis of many sicknesses. Despite being categorised as a whole category (in our experiments), some blood anoma-

lies (e.g. iron deficiency) are included in the artefact class. The proposed classification scheme can be extended to identify these conditions as well.

10.2 Improvements and Extensions

The possible improvements and extensions for the proposed methods are listed below.

- The minimum area watershed transformation (described in Section 4.4) can be extended to work with other attributes. Thus, it may be possible to provide a generalised attribute watershed transform.

- It is possible to evaluate the colour normalisation method (described in Section 5.2) by acquiring the same field of a slide with different acquisition settings and comparing the difference (e.g. RMS) of the outputs.

- The stained-pixel detection method (described in Section 6.2) can be improved by enhancing the feature and classification scheme (see the summary in the previous section). Thus, it may be possible to bypass the morphological operations, which rely on the average cell area estimate by area granulometry. Therefore, the method would be applicable to thick film analysis or other blood parasites.

- For the diagnosis tasks KNN classification algorithm was used with simple distance metrics. Some adaptive distance measures are known to improve the KNN algorithm's performance on some common data sets [163]. This can be investigated further. The adaptive distance measures usually incorporate more sophisticated and complex learning algorithms (e.g. support vector machine) to KNN. However, it should be taken into account that the simplicity of KNN provides a good flexibility for possible future extensions to the system, e.g. for diagnosis of other parasites.

- The same feature set was used for parasite detection, species and life-stage recognition tasks. Considering the problems are different, feature relevancies should

have been investigated for each task and each classification scheme separately. This issue can be investigated further when a significant number of samples (i.e. smears, slides) are available.

- In manual microscopy examination, the species and life-stage recognition tasks can be performed contextually for a whole slide. Therefore, the appearance of one parasite can be used as a clue for another parasite. This issue can be investigated further when a significant number of samples (i.e. smears, slides) are available.

10.3 Conclusions for CDM and Future Directions

The set of solutions proposed here provided a well-defined basis for realising the computerised diagnosis of malaria. It is possible to design a generalised diagnosis tool covering other peripheral blood parasites since the methodology does not depend on the characteristics of *Plasmodium*, or its appearance in a particular set of images. This is very important because a diagnosis tool would be more useful if other parasites can be recognised too.

Microscopy examination of the blood film currently is still "the gold standard" for malaria diagnosis despite being invented in the late 19th century. The manual microscopy requires proper training and has the disadvantage of being subjective.

The most sensitive technique to detect malaria parasites is PCR (Polymerase Chain Reaction) [17]. However, PCR is not suitable for diagnosis due to its complexity and slowness. Emerging new technologies such as Rapid Diagnostic Tests (RDT) do not require any special equipment and training. Their detection sensitivity is lower but comparable to manual microscopy. However, they do not provide quantification of the results, species discrimination (for all species) [11], and life-stage identification.

Computerised Diagnosis of Malaria (CDM) aims to perform automated microscopy diagnosis. Its per-slide detection sensitivity may not be expected to be better than an *expert* microscopist but a result equivalent would be acceptable. However, CDM

has certain advantages such as being objective, improvable, and being cheaper when the personnel training costs are considered. As a microscopy diagnosis tool, it requires stained blood samples (i.e. smear) and a digital camera mounted microscope, with an additional computer. The preparation of a single slide takes approximately 30-40 minutes. This can be done in batches, however, a potential which may worth investigating is whether the diagnosis can be performed (i.e. parasite detection) without staining.

In this work, due to limited number of samples/slides, per-slide diagnosis tests could not be performed. A large-scale per-slide diagnosis test is necessary to reveal the actual diagnosis performance of the overall method. This test should include a significant number of test smears (slides) (both positive and negative). The positive smears (i.e. infected) should include all four species and all life-stages. The results should be compared to manual expert microscopy and if possible to RDT results. This is the primary future direction of this study.

It should also be noted that the thick film examination sensitivity in microscopy diagnosis is approximately 10 times higher than for thin films [170]. However, species recognition is more difficult due to deformed parasites and being performed on thin blood films. If the CDM detection sensitivity in thin blood films is found satisfactory in large-scale tests, thick blood film analysis may not be essential. However, if it is unsatisfactory thick film analysis should be investigated. Some of the tools developed in this thesis can be applied to thick films with minor modifications.

In order to be practical the computerised malaria diagnosis system must be provided with automatic slide positioning and image capture facilities. As performing diagnosis of a single sample, the slide must be re-positioned at least 100 times, focused, and captured. The system would be highly impractical if manual assistance was required. Some of the state-of-art microscopes located in well-equipped laboratories already have these functions. However, one of the general aims of this study was also to produce a cost effective diagnosis system that can be used especially in the economically weaker

areas where malaria is endemic and causing a serious number of deaths. The UoW microscope-camera setup that was built in this study is comprised of a low cost microscope and a consumer camera. However, it requires manual positioning of the slides and manual capturing of the images. A possible solution to this problem can be a motorised optical imaging system, which is customised for computerised diagnosis. The costs may be significantly reduced since a customised system will not require many of the general-purpose functions and adjustments of both microscope and camera.

REFERENCES

[1] E. Korenromp, J. Miller, B. Nahlen, T. Wardlaw, and M. Young, "World malaria report 2005," World Health Organization, Geneva, Tech. Rep., 2005.

[2] J. Gallup and J. Sachs, "The economic burden of malaria," *Journal of Tropical Medicine*, vol. 64 (Suppl 1-2), pp. 85–96, 2001.

[3] Roll-Back-Malaria, "http://www.rollbackmalaria.org/," Date Accessed, June 2006.

[4] "What is malaria? fact st no 94," Date Accessed, June 2006. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs094/en

[5] "Laboratory identification of parasites of public health concern," Centers for Disease Control Prevention, National Center for Infectious Diseases, Division of Parasitic Diseases, Date Accessed, June 2006. [Online]. Available: http://www.dpd.cdc.gov/dpdx/HTML/Malaria.htm

[6] "Malaria journal," Date Accessed, May 2006. [Online]. Available: http://www.malariajournal.com/

[7] G. Coatney, W. Collins, M. Warren, and P. Contacos, *The Primate Malarias*. U.S. Department of Health, Education and Welfare, 1971.

[8] L. J. Bruce-Chwatt, *Essential Malariology*. Hodder Arnold, 2002.

[9] WHO, *Basic malaria microscopy Part I. Learner's Guide*. World Health Organization, 1991.

[10] "Malaria: An on-line resource," Division of Laboratory Medicine at Royal Perth Hospital, Date Accessed Jan 2007. [Online]. Available: http://-www.rph.wa.gov.au/malaria.html

[11] WHO, "New perspectives, malaria diagnosis," World Health Organization, Geneva, Tech. Rep., 2000.

[12] R. E. Coleman, N. Maneechai, N. Rachaphaew, C. Kumpitak, R. Miller, V. Soyseng, K. Thimasarn, and J. Sattabongkot, "Comparison of field and expert laboratory microscopy for active surveillance for asymptomatic plasmodium falciparum and Plasmodium Vivax in Western Thailand," *American Journal of Tropical Medicine and Hygiene*, vol. 67, pp. 141–144, 2002.

[13] K. Mitiku, G. Mengistu, and B. Gelaw, "The reliability of blood film examination for malaria at the peripheral health unit," *Ethiopian Journal of Health Development*, vol. 17, no. 3, pp. 197–204, 2003.

[14] I. Bates, V. Bekoe, and A. Asamoa-Adu, "Improving the accuracy of malaria-related laboratory tests in ghana," *Malaria Journal*, vol. 3, p. 38, 2004.

[15] C. K. Murray, D. Bell, R. A. Gasser, and C. Wongsrichanalai, "Rapid diagnostic testing for malaria," *Tropical Medicine and International Health*, vol. 8, pp. 876–883, 2003.

[16] G. Snounou, S. Viriyakosol, X. Zhu, W. Jarra, L. Pinheiro, V. do Rosario, S. Thaithong, and K. Brown, "High sensitivity of detection of human malaria parasites by the use of nested polymerase chain reaction," *Molecular Biochemistry Parasitology*, vol. 61, no. 2, pp. 315–320, 1993.

[17] A. Moody, "Rapid diagnostic tests for malaria parasites," *Clinical Microbiology Reviews*, vol. 15, pp. 66–78, 2002.

[18] R. E. Coleman, J. Sattabongkot, S. Promstaporm, N. Maneechai, B. Tippayachai, A. Kengluecha, N. Rachapaew, G. Zollner, R. S. Miller, J. A. Vaughan, K. Thimasarn, and B. Khuntirat, "Comparison of PCR and microscopy for the detection of asymptomatic malaria in a Plasmodium falciparum/vivax endemic area in Thailand," *Malaria Journal*, vol. 5, p. 121, 2006.

[19] N. Hemvani, D. Chitnis, D. Dixit, and M. Asolkar, "Acridine orange stained blood wet mounts for fluorescent detection of malaria," *Indian Journal of pathology and Microbiology*, vol. 42, pp. 125–128, 1999.

[20] Y. Htut, K. Aye, K. Han, M. Kyaw, K. Shimono, and S. Okada, "Feasibility and limitations of acridine orange fluorescence technique using a malaria diagnosis microscope in Myanmar," *Acta Medica Okayama*, vol. 56, no. 5, pp. 219–22, 2002.

[21] T. Hanscheid, "Current strategies to avoid misdiagnosis of malaria," *Clinical Microbiology and Infection*, vol. 9, no. 6, pp. 497–504, 2003.

[22] T. Hanscheid, "Diagnosis of malaria: a review of alternatives to conventional microscopy," *Int. Journal of Laboratory Hematology*, vol. 21, pp. 235–245, 1999.

[23] M. T. Makler, C. J. Palmer, and A. L. Ager, "A review of practical techniques for the diagnosis of malaria," *Annals of Tropical Medicine & Parasitology*, vol. 92, no. 4, pp. 419–433, 1998.

[24] L. M. Milne, M. S. Kyi, P. L. Chiodini, and D. C. Warhurst, "Accuracy of routine laboratory diagnosis of malaria in the United Kingdom," *Journal of Clinical Pathology*, vol. 47(8), pp. 740–742, 1994.

[25] WHO, "Guidelines for the treatment of malaria," World Health Organisation, Geneva, Tech. Rep., 2006.

[26] J. Angulo and G. Flandrin, "Automated detection of working area of peripheral blood smears using mathematical morphology," *Analytical Cellular Pathology*, vol. 25, no. 1, pp. 37–49, 2003.

[27] L. Garcia, *Diagnostic Medical Parasitology 4th Ed.* Herndon, USA: ASM Press, 2001.

[28] E. R. Dougherty and R. A. Lotufo, *Hands-on Morphological Image Processing.* USA: SPIE Press, 2003.

[29] P. Soille, *Morphological Image Analysis.* Heidelberg, Germany: Springer-Verlag, 2003.

[30] R. Davies, *Machine Vision: theory, algorithms, practicalities 2nd Ed.* London, UK: Academic Press, 1997.

[31] A. Webb, *Statistical pattern recognition 2nd Ed.* New York, USA: J Wiley and Sons Inc., 2002.

[32] C. M. Bishop, *Neural Networks for Pattern Recognition.* UK: Oxford University Press, 2004.

[33] R. Duda, P. Hart, and D. Stork, *Pattern Classification.* New York, USA: Wiley-Interscience Publication, 2000.

[34] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Analysis of infected blood cell images using morphological operators," *Image and Vision Computing*, vol. 20, no. 2, pp. 133–146, February 2002.

[35] K. N. R. M. Rao, "Application of mathematical morphology to biomedical image processing," Ph.D. dissertation, University of Westminster, London, UK, 2004.

[36] F. B. Tek, A. Dempster, and I. Kale, "Malaria parasite detection in peripheral blood images," in *Proc. Medical Image Understanding and Analysis Conference*, Manchester, UK, 2006.

[37] F. B. Tek, A. Dempster, and I. Kale, "Malaria parasite detection in peripheral blood images," in *Proc. British Machine Vision Conference*, Edinburgh, UK, 2006.

[38] F. B. Tek, A. G. Dempster, and I. Kale, "Noise sensitivity of watershed segmentation for different connectivity: experimental study," *IEE Short Letters*, vol. 40, no. 21, pp. 1332–1333, 2004.

[39] F. B. Tek, A. G. Dempster, and I. Kale, "A colour normalization method for giemsa-stained blood cell images," in *Proc. Signal Processing and Applications (IEEE), SIU2006*, Antalya, Turkey, 2006.

[40] F. B. Tek, A. G. Dempster, and I. Kale, "Blood cell segmentation using minimum area watershed and circle radon transformations," in *Proc. Int. Symp. on Mathematical Morphology*, Paris, France, 2005.

[41] J. Serra, *Image Analysis and Mathematical Morphology.* Philadelphia, USA: Academic Press Inc, 1983.

[42] L. Vincent, "Morphological area openings and closings for greyscale images," in *Proc. NATO Shape in Picture Workshop*, Driebergen, The Netherlands, 1992.

[43] L. Vincent, "Granulometries and opening trees," *Fundamenta Informaticae*, vol. 41, pp. 57–90, 2000.

[44] P. Maragos, "Pattern spectrum and multiscale shape respresentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 2, pp. 701–716, 1989.

[45] K. N. R. M. Rao and A. Dempster, "Area-granulometry: an improved estimator of size distribution of image objects," *IEE Electronics Letters*, vol. 37, pp. 50–951, 2001.

[46] L. Vincent, "Morphological grayscale reconstruction in image analysis applications and efficient algorithms," *IEEE Trans. on Image Processing*, vol. 2, pp. 176–201, 1993.

[47] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 583–598, 1991.

[48] P. A. Flach, "The many faces of ROC analysis in machine learning," in *ICML Tutorial*, 2004. [Online]. Available: http://www.cs.bris.ac.uk/flach/-ICML04tutorial/index

[49] P. Flach, "The geometry of ROC space: understanding machine learning metrics through ROC isometrics," in *Proc. Int. Conf. on Machine Learning*, Washington DC, USA, 2003.

[50] P. J. Edwards, A. P. King, C. R. Maurer, D. A. de Cunha, D. J. Hawkes, D. L. Hill, R. P. Gaston, M. R. Fenlon, A. Jusczyzck, A. J. Strong, C. L. Chandler, and M. J. Gleeson, "Design and evaluation of a system for microscope-assisted guided interventions," *IEEE Trans. On Medical Imaging*, vol. 19, no. 11, pp. 1082–1093, 2000.

[51] Q. Guo, M. Ohsuga, and K. Milroi, "Generation of high-quality images for telemedicine and tele-pathology efforts," in *Proc. 20th Ann. Int. Conf. IEEE EMBS*, NY, USA, 1998.

[52] U. Catalyurek, M. Beynon, C. Chang, T. Kurc, A. Sussman, and J. Saltz, "The virtual microscope," *IEEE Trans. on Information Technology in Biomedicine*, vol. 7, pp. 230–248, 2003.

[53] J. Schmitz, O. Bollmann, C. Vogel, and R. Bollmann, "Virtual microscopy (remote patchwork) as a new technique for tele-consultation and tele-education," *Electronic Journal of Pathology and Histology*, vol. 9.2, pp. 32–005, 2003.

[54] A. Parra, J. Puentes, and A. Herrera, "Remote microscope control for cervical cancer telediagnostic," in *Proc. 25th Ann. Int. Conf. IEEE EMBS*, Cancun, Mexico, 2003.

[55] A. Wetzel, P. Feineigle, and J. Gilbertson, "Design of a high-speed slide imaging system for pathology," in *Proc. IEEE Int. Symp. on Biomedical Imaging*, Washington DC, USA, 2002.

[56] Y. Yagi and J. R. Gilbertson, "Digital imaging in pathology: the case for standardization," *Journal of Telemedicine and Telecare*, vol. 11, pp. 109–116, 2005.

[57] Y.-C. Chang and J. Reid, "RGB calibration for color image analysis in machine vision," *IEEE Trans. on Image Processing*, vol. 5, pp. 1414–1422, 1996.

[58] K. Barnard, "Practical colour constancy," Ph.D. thesis, Simon Fraser University School of Computing Science, 1999. [Online]. Available: ftp://fas.sfu.ca/pub/cs/theses/1999/KobusBarnardPhD.ps.gz

[59] H.-Z. Zhang, K.-Q. Wang, X.-S. Jin, and D. Zhang, "SVR based color calibration for tongue image," in *Proc. Int. Conf. on Machine Learning and Cybernetics*, Guangzhou, China, 2005.

[60] Y. Haeghen, J. Naeyaert, I. Lemahieu, and W. Philips, "An imaging system with calibrated color image acquisition for use in dermatology," *IEEE Trans. on Medical Imaging*, vol. 19, pp. 722–730, 2000.

[61] C. Grana, G. Pellacani, and S. Seidenari, "Practical color calibration for dermoscopy, applied to a digital epiluminescence microscope," *Skin Research and Technology*, vol. 11, pp. 242–247, 2005.

[62] C. M'unzenmayer, D. Paulus, and T. Wittenberg, "A spectral color correction framework for medical applications," *IEEE Trans. On Biomedical Engineering*, vol. 53, pp. 254–265, 2006.

[63] H.-C. Lee, *Introduction to Color Imaging Science*. UK: Cambridge University Press, 2005.

[64] T. Abe, M. Yamaguchi, Y. Murakami, N. Ohyama, and Y. Yagi, "Color correction of pathological images for different staining-condition slides," in *Proc. 6th International HealthCom Workshop*, Odawara, Japan, 2004.

[65] K. Fujii, M. Yamaguchi, N. Ohyama, and K. Mukai, "Development of support systems for pathology using spectral transmittance: the quantification method of stain conditions," in *Proc. SPIE Int. Soc. Opt. Eng.*, San Jose, California, 2002.

[66] D. Comaniciu, D. Foran, and P. Meer, "Shape-based image indexing and retrieval for diagnostic pathology," in *Proc. 14 th Int. Conference on Pattern Recognition*, Brisbane, 1998.

[67] N. Sinha and A. G. Ramakrishnan, "Blood cell segmentation using EM algorithm," in *Proc. Indian Conf. on Computer Vision, Graphics and Image Processing*, Ahmadabad, India, 2002.

[68] K. Jiang, Q. Liao, and S. Dai, "A novel white blood cell segmentation scheme using scale-space filtering and watershed clustering," in *Proc. 2nd Int. Conf. on Machine Learning and Cybernetics*, Beijing, China, 2003.

[69] J. H. Carr and B. Rodak, *Clinical Hematology Atlas*. Philadelphia, US: W. B. Saunders, 1998.

[70] M. R. Howard and P. Hamilton, *An Illustrated Colour Text: Haematology 2nd Ed.* Oxford, UK: Churchill Livingstone, 2002.

[71] "Cell atlas, www.cellatlas.com," internet resource, Date Accessed, June, 2006.

[72] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra, "Segmentation of blood images using morphological operators," in *Proc. Int. Conf. on Pattern Recognition*, Barcelona, Spain, 2000.

[73] C. Di Ruberto, A. G. Dempster, S. Khan, and B. Jarra, "Automatic thresholding of infected blood images using granulometry and regional extrema," in *Proc. Int. Conf. on Pattern Recognition*, Barcelona, Spain, 2000.

[74] N. E. Ross, C. J. Pritchard, D. M. Rubin, and A. G. Duse, "Automated image processing method for the diagnosis and classification of malaria on thin blood smears," *Medical Biolology Enginerring Computing*, vol. 44, pp. 427–436, 2006.

[75] E. Breen and R. Jones, "Attribute openings, thinnings, and granulometries," *Computer Vision and Image Understanding*, vol. 64, pp. 377–389, 1996.

[76] P. Salembier, A. Oliveras, and L. Garrido, "Anti-extensive connected operators for image and sequence processing," *IEEE Trans. on Image Processing*, vol. 7, pp. 555–570, 1998.

[77] E. R. Urbach and M. H. F. Wilkinson, "Shape-only granulometries and grey-scale shape filters," in *Proc. Int. Symp. on Mathematical Morphology*, New South Wales, Australia, 2002.

[78] E. R. Urbach, J. B. T. M. Roerdink, and M. H. F. Wilkinson, "Connected rotation-invariant size-shape granulometries," in *Proc. Int. Conf. Pattern Recognition*, Cambridge, UK, 2004.

[79] E. R. Urbach, N. J. Boersma, and M. H. F. Wilkinson, "Vector-attribute filters," in *Proc. Int. Symp. Mathematical Morphology*, Paris, France, 2005.

[80] A. Meijster and M. Wilkinson, "Fast computation of morphological area pattern spectra," in *Proc. Int. Conf. on Image Processing*, 2001.

[81] I. Ceske, "A fast segmentation scheme for white blood cell images," in *Proc. 11th IAPR International Conference on Pattern Recognition*, The Hague, The Netherlands, 1992.

[82] P. Zhou and D. Pycock, "Robust statistical models for cell image interpretation," *Image and Vision Computing*, vol. 15, pp. 307–316, 1997.

[83] S. Bikhet, A. Darwish, H. Tolba, and S. Shaheen, "Segmentation and classification of white blood cells," in *Proc. IEEE ICASSP*, Istanbul, Turkey, 2000.

[84] Q. Liao and Y. Deng, "An accurate segmentation method for white blood cell images," in *Proc. IEEE Int. Symp. on Biomedical Imaging*, Washington DC, USA, 2002.

[85] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation," *Journal Of Electronic Imaging*, vol. 13, pp. 146–168, 2004.

[86] N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 9, pp. 62–66, 1979.

[87] M. Kass, A. Witkin, and D. Terzopoulos, "Snake: Active contour models," *Int. Journal of Computer Vision*, vol. 1, pp. 321–331, 1987.

[88] J. Park and J. M. Keller, "Snakes on the watershed," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1201–1205, 2001.

[89] M. A. Roula, A. Bouridune, and F. Kurugollu, "An evolutionary snake algorithm for the segmentation of nuclei in histopathological images," in *Proc. Int. Conf. on Image Processing*, Singapore, 2004.

[90] G. Ongun, U. Halici, K. Leblecioglu, V. Atalay, M. Beksac, and S. Beksac, "An automated differential blood count system," in *Proc. 23rd Ann. Int. Conf. IEEE EMBS*, Istanbul, Turkey, 2001.

[91] C. Zimmer, E.Labruyere, V. Meas-Yedid, N. Guillen, and J.-C. Olivo-Marin, "Quantitative analysis of cellular shape and motility with improved active contours," in *Proc. IEEE Int. Symp. on Biomedical Imaging*, Washington DC, USA, 2002.

[92] C. Walhby, "Algorithms for applied digital image cytometry," Ph.D. dissertation, Uppsala University, Faculty of Science and Technology, Sweden, 2003.

[93] M. D. Andrade, G. Bertrand, and A. Araujo, "Segmentation of microscopic images by flooding simulation: A catchment basins merging algorithm," in *Proc. SPIE Int. Soc. Opt. Eng.* San Jose, CA, USA: SPIE, 1997.

[94] A. Cosio, F. M. Flores, J. A. P. Castaneda, M. A. Solano, and S. Tato, "Automatic counting of immunocytochemically stained cells," in *Proc. 25th Ann. Int. Conf. IEEE EMBS*, Cancun,Mexico, 2003.

[95] K. N. R. M. Rao and A. Dempster, "Modification on distance transform to avoid over-segmentation and under-segmentation," in *Proc. Int. Symp. on Video/Image Processing and Multimedia Communications*, Crotia, 2002.

[96] M. van Ginkel, C. L. Hendriks, and L. van Vliet, "A short introduction to the radon and hough transforms and how they relate to each other," Quantitative Imaging Group, Delft University of Technology, Netherlands, Tech. Rep., 1997.

[97] L. Wei, X. Jianhong, and E. M.-Tzanakou, "A computational intelligence system for cell classification," in *Proc. IEEE Information Technology Applications in Biomedicine*, Washington DC, USA, 1998.

[98] E. M. Tzanakou, H. Sheikh, and B. Zhu, "Neural networks and blood cell identification," *Journal of Medical Systems*, vol. 21, pp. 201–210, 1997.

[99] E. Harth and E. Tzanakou, "ALOPEX: A stochastic method for determining visual receptive fields," *Vision Research*, vol. 14, pp. 1475–1482, 1974.

[100] S. Sanei and T. Lee, "Bayesian classification of eigencells," in *Proc. Int. Conf. on Image Processing*, Rochester, NY, 2002.

[101] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71–86, 1991.

[102] D. Comaniciu and P. Meer, "Cell image segmentation for diagnostic pathology," in *Advanced Algorithmic Approaches to Medical Image Segmentation: State-Of-The-Art Applications in Cardiology, Neurology, Mammography and Pathology.* Springer, 2001.

[103] G. Ongun, U. Halici, K. Leblebicioglu, V. Atalay, M. Beksac, and S. Beksac, "Feature extraction and classification of blood cells for an automated differential blood count system," in *Proc. Int. Joint Conf. on Neural Networks*, Washington DC, USA, 2001.

[104] V. Kovalev, A. Grigoriev, and A. Hyo-Sok, "Robust recognition of white blood cell images," in *Proc. 13th Int. Conf. on Pattern Recognition*, Vienna, Austria, 1996.

[105] V. Piuri and F. Scotti, "Morphological classification of blood leucocytes by microscope images," in *Proc. IEEE Int. Conf. on Computational Intelligence for Measurement Systems and Applications*, Boston, USA, 2004.

[106] D. Sabino, L. Costa, S. Martins, R. Calado, and M. Zago, "Automatic leukemia diagnosis," *Acta Microscopica*, vol. 12, no. 1, pp. 1–6, 2003.

[107] A. R. J. Katz, "Image analysis and supervised learning in the automated differentiation of white blood cells from microscopic images," Master's thesis, Department of Computer Science, RMIT, 2000.

[108] D. M. Ushizima, A. C. Lorena, and A. Carvalho, "Support vector machines applied to white blood cell recognition," in *Proc. Int. Conf. on Hybrid Intelligent Systems*, Rio de Janeiro - Brazil, 2005.

[109] K. N. R. M. Rao, A. G. Dempster, B. Jarra, and S. Khan, "Automatic scanning of malaria infected blood slide images using mathematical morphology," in *IEE Seminar Medical Applications of Signal Processing*, London, UK, 2002.

[110] S. W. Sio, W. Sun, and S. K. *et.al*, "Malariacount: An image analysis-based program for the accurate determination of parasitemia," *Journal of Microbiological Methods*, vol. 68, pp. 11–18, 2007.

[111] M. Kettlehut and C. Rogers, "Update in blood and tissue parasitology," Biomedical Short Course Unit, University of Westminster, Tech. Rep., 2005.

[112] TWAIN Working Group, "TWAIN specification (version 1.9)," 2000, Date Accessed Aug 2006. [Online]. Available: http://www.twain.org

[113] M. Spencer, *Fundamentals of Light Microscopy.* UK: Cambridge University Press, 1982.

[114] S. Volkova, N. Ilyasova, A. Ustinov, and A. Khramov, "Methods for analysing the images of blood preparations," *Optics & Laser Technology*, vol. 27, no. 2, pp. 255–261, 1995.

[115] F. Meyer and S. Beucher, "Morphological segmentation," *J. Vis. Comms. Image Representation*, vol. 1, pp. 21–46, 1990.

[116] L. Najman and M. Schmitt, "Geodesic saliency of watershed contours and hierarchical segmentation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 18, pp. 1163–1173, 1996.

[117] O. Lezoray and H. Cardot, "Bayesian marker extraction for color watershed in segmenting microscopic images," in *Proc. 16th Int. Conf. Pattern Recognition*, Quebec, Canada, 2002.

[118] C. Vachier and F. Meyer, "Extinction value: a new measurement of persistence," in *Proc. IEEE Workshop on Nonlinear Signal and Image Processing*, Greece, 1995.

[119] R. A. Mclaughlin, "Randomized hough transform: Improved ellipse detection with comparison," *Pattern Recognition Letters*, vol. 19, no. 3-4, pp. 299–305, 1998.

[120] E. Davies, "Finding ellipses using the generalised hough transform," *Pattern Recognition Letters*, vol. 9, pp. 87–96, 1989.

[121] N. Guil and E. L. Zapata, "Lower order circle and ellipse hough transform," *Pattern Recognition*, vol. 30, pp. 1729–1744, 1997.

[122] J. Illingworth and J. Kittler, "A survey of the hough transform," *Computer Vis. Graph. and Img. Processing*, vol. 44, pp. 87–116, 1988.

[123] R. O. Duda and P. E. Hart, "Use of the hough transformation to detect lines & curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[124] V. F. Leavers, "Use of the two-dimensional radon transform to generate a taxonomy of shape for the characterization of abrasive powder particles," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1411–1423, 2000.

[125] K. V. Hansen and P. A. Toft, "Fast curve estimation using preconditioned generalized radon transform," *IEEE Trans. on Image Processing*, vol. 5, no. 12, pp. 1651–1661, 1996.

[126] C. L. Hendriks, M. van Ginkel, P. Verbeek, and L. van Vliet, "The generalised radon transform: Sampling and memory considerations," in *Proc. of Computer Analysis of Images and Patterns*. Springer Verlag, 2003.

[127] C. Di Ruberto, A. G. Dempster, S. Khan, and B. Jarra, "Morphological image processing for evaluating malaria disease," in *Proc. Int. Workshop on Visual Form*, Capri, Italy, 2001.

[128] S. Hordley and G. Finlayson, "Re-evaluating colour constancy algorithms," in *Proc. of the 17th International Conference on Pattern Recognition, Cambridge, UK*, 2004.

[129] K. Barnard, L. Martin, A. Coath, and B. Funt, "A comparison of computational color constancy algorithms. ii. experiments with image data," *IEEE Transactions on Image Processing*, vol. 11, pp. 985–996, 2002.

[130] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *Int. Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[131] R. Bellman, *Adaptive Control Processes.* Princeton University Press, NJ, 1961.

[132] S. L. Phung, A. Bouzerdoum, and D. Chai, "Skin segmentation using color pixel classification: analysis and comparison," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 148–154, 2005.

[133] B. Zadrozny and C. Elkan, "Obtaining calibrated probability estimates from decision trees and naive bayesian classifiers," in *Proc. Int. Conf. on Machine Learning*, Massachusetts, USA, 2001.

[134] I. Cohen and M. Goldszmidt, "Properties and benefits of calibrated classifiers," in *Proc. 15th European Conf. on Machine Learning*, Pisa, Italy, 2004.

[135] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119–1125, 1994.

[136] J. Huang, S. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Spatial color indexing and applications," *Int. Journal of Computer Vision*, vol. 35, no. 3, pp. 245–268, 1999.

[137] P. Chang and J. Krumm, "Object recognition with color cooccurence histograms," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, Fort Collins, CO, USA, 1999.

[138] X. Tang, K. Stewart, L. Vincent, H. Huang, M. Marra, S. Gallager, and C. Davis, "Automatic plankton image recognition," *Artificial Intelligence Review*, vol. 12, pp. 177–99, 1998.

[139] P.-F. Yang and P. Maragos, "Morphological systems for character image processing and recognition." in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Minneapolis, USA, 1993.

[140] Y. Chen and E. Dougherty, "Texture classification by gray-scale morphological granulometries," in *Proc. SPIE Visual Communications and Image Processing*, Boston, USA, 1992.

[141] M.-K. Hu, "Visual pattern recognition by moment invariants," *IEEE Trans. on Information Theory*, vol. 8, pp. 179–187, 1962.

[142] X. W. (ed. James Arvo), "Colour indexing: Efficient statistical computation for optimal color quantization," in *Graphics Gems II.* Philadelphia, USA: Academic Press, 1991.

[143] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: analysis and implementation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–891, 2002.

[144] P. M. Narendra and K. Fukunaga, "A branch and bound algorithm for feature subset selection," *IEEE Trans. on Computers*, vol. 26, pp. 917–922, 1977.

[145] P. Somol, P. Pudil, and J. Kittler, "Fast branch & bound algorithms for optimal feature selection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 900–912, 2004.

[146] J. Kittler, "Feature set search algorithms," in *Proc. Pattern Recognition and Signal Processing*, The Netherlands, 41-60,1978.

[147] A. Jain and D. Zongker, "Feature selection: Evaluation, application,and small sample performance," *IEEE Trans. On Pattern Analysis And Machine Intelligence*, vol. 19, pp. 153–158, 1997.

[148] G. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in *Proc. 11th Int. Conf. on Machine Learning*, New York, USA, 1994.

[149] P. Krizek, J. Kittler, and V. Hlavac, "Feature selection based on the training set manipulation," in *Proc. The 18th Int. Conf. on Pattern Recognition*, Hong Kong, 2006.

[150] G. Xuan, X. Zhu, P. Chai, Z. Zhang, Y. Q. Shi, and D. Fu, "Feature selection based on the Bhattacharyya distance," in *Proc. Int. Conf. on Pattern Recognition*, Hong Kong, China, 2006.

[151] S. Raudys, "Accuracy of feature selection and extraction in statistical and neural network pattern classification," in *Proc. 11th IAPR Pattern Recognition*, The Hague, The Netherlands, 1992.

[152] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Trans. on Information Theory*, vol. IT-13, pp. 21–27, 1967.

[153] K. Fukunage and P. Narendra, "A branch and bound algorithm for computing k-nearest neighbors," *IEEE Trans. on Computers*, vol. C-24, pp. 750–753, 1975.

[154] J. McNames, "A fast nearest-neighbor algorithm based on a principal axis search tree," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 964–976, 2001.

[155] B. Zhang and S. Srihari, "Fast k-nearest neighbor classification using cluster-based trees," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 525–528, 2004.

[156] D. Wettschereck and T. G. Dietterich, "Locally adaptive nearest neighbor algorithms," in *Proc. Advances in Neural Information Processing Systems*, Denver, USA, 1994.

[157] C. Domeniconi, D. Gunopulos, and J. Peng, "Large margin nearest neighbor classifiers," *IEEE Trans. on Neural Networks*, vol. 16, pp. 899–910, 2005.

[158] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth movers distance as a metric for image retrieval," *Int. Journal Of Computer Vision*, vol. 40, pp. 99–121, 2000.

[159] H. Ling and K. Okada, "Diffusion distance for histogram comparison," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, New York, USA, 2006.

[160] T. Hastie and R. Tibshirani, "Discriminant adaptive nearest neighbor classification and regression," in *Advances in Neural Information Processing Systems*, vol. 8. Denver, USA: The MIT Press, 1996.

[161] J. Peng, D. R. Heisterkamp, and H. K. Dai, "LDA/SVM driven nearest neighbor classification," *IEEE Trans. on Neural Networks*, vol. 14, pp. 940–942, 2003.

[162] S. Geva and J. Sitte, "Adaptive nearest neighbor pattern classification," *IEEE Trans. on Neural Networks*, vol. 2002, pp. 318–322, 1991.

[163] C. Domeniconi, J. Peng, and D. Gunopulos, "Locally adaptive metric nearest neighbor classification," *IEEE Trans. Pattern Analysis Machine Intelligence*, vol. 24, pp. 1281–1285, 2002.

[164] T.-W. Loong, "Understanding sensitivity and specificity with the right side of the brain," *BMJ*, vol. 327, pp. 716–719, 2003.

[165] K. Fukunaga and R. Hayes, "Estimation of classifier performance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 1087 – 1101, 1989.

[166] S. Aksoy and R. M. Haralick, "Feature normalization and likelihood-based similarity measures for image retrieval," *Pattern Recognition Letters*, vol. 22, pp. 563–582, 2001.

[167] "Int. conf. on machine learning workshop on learning from imbalanced data sets (ii)," Internet, August 2003. [Online]. Available: http://www.site.uottawa-.ca/ nat/Workshop2003/workshop2003.html

[168] M. A. Maloof, P. Langley, T. O. Binford, R. Nevatia, and S. Sage, "Improved rooftop detection in aerial images with machine learning," *Machine Learning*, vol. 53, pp. 157–191, 2003.

[169] V. Franc and V. Hlavac, "Statistical Pattern Recognition Toolbox for Matlab, ctucmp200408," Czech Technical University in Prague, Tech. Rep., 2004.

[170] D. C. Warhurst and J. E. Williams, "Laboratory diagnosis of malaria," *Journal of clinical pathology*, vol. 49, pp. 533–538, 1996.

[171] J. D. Maguire, E. R. .Lederman, M. J. Barcus, W. A. P. O'Meara, R. G. Jordon, S. Duong, S. Muth, P. Sismadi, M. J. Bangs, W. RoyPrescot, J. K. Baird, and C. Wongsrichanalai, "Production and validation of durable, high quality standardized malaria microscopy slides for teaching, testing and quality assurance during an era of declining diagnostic proficiency," *Malaria Journal*, vol. 5, p. 92, 2006.